HISTORICAL AND CONTEMPORARY PROCESSES SHAPING POPULATION GENETIC STRUCTURE IN AN ANADROMOUS FISH (*OSMERUS MORDAX*)

by

Mark W. Coulson

Submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
February 2014

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The spatial scale at which populations are genetically structured is of immense interest for the understanding of a species' ecology and evolutionary biology. This can have important implications for management of resources as well as predicting responses to future change. Rainbow smelt (*Osmerus mordax*) is an anadromous species with a relatively short freshwater residence time compared to other species with similar life-history strategies. Therefore, while they offer the opportunity to sample distinct spawning aggregations, they also offer an insight into the relative roles of contemporary and historical factors shaping connectivity among marine populations, an area of great interest, and for which further understanding is required. With the use of both mitochondrial DNA and nuclear microsatellite markers, I explored the historical and contemporary factors influencing population structure in smelt. While previous phylogeographic work on this species has resolved two mtDNA lineages dating back to previous glacial episodes, I document the discovery of a zone of contact between these lineages in Newfoundland. This is in addition to the established contact zone in the St. Lawrence estuary, and results in a longitudinal distribution of the races with one race predominating on opposite ends of the species distribution, while the other race is geographically intermediate. Patterns of nuclear genetic variation largely mirror the phylogeographic signals in Newfoundland and suggest a more recent colonization of the Avalon Peninsula as well as implicating a remnant historical signal of colonization of the west coast of Newfoundland from the mainland. In addition, contrasting patterns of genetic diversity and levels of differentiation were apparent between the mainland and Newfoundland and suggest differing scales of dispersal within this species. While the population structure within Newfoundland is most consistent with dispersal restricted to within bays, larger scale biogeographic regions were identified in the mainland range, suggesting dispersal is more common and widespread. In addition, sampling of different run times (i.e. 'early' vs. 'late') demonstrated the potential for isolation by time when spawning events are separated by a break in activity. Overall, these results shed light into the possible roles of both historical and contemporary factors shaping the dynamics and connectivity among populations.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor, Dr. Paul Bentzen. Paul you have always been available for discussion, comments and bouncing off of ideas. I have appreciated all of your guidance, support and encouragement over the course of this thesis. I realize this has taken much longer than either of us had anticipated, and I am grateful for your patience and help in getting me to this point.

To Dr. Daniel Ruzzante, my thanks for your service on my thesis committee and for your participation in many discussions over the years. You have helped to ensure that there is a continuity and flow connecting the different aspects of this work and keeping expectations in check regarding how much one can accomplish.

To Dr. Rod Bradford, also my thanks to you for your contribution to my committee. Rod, you ensured that I didn't get so focused on the genetic results as to lose sight of the wider relevant species' biology and ecology. You have helped to foster my critical thinking and interpretation skills and your attention to detail as to how thoughts are worded and expressed has been a very valuable lesson.

To Ian Paterson, many thanks for your technical advice and assistance over the years. To the many students, postdocs and technicians with whom I shared the lab, many thanks to you all for your exchange of ideas, assistance with lab and fieldwork and reminder to stop and have a bit of fun every now and again. There are too many of you to name, so I hope you know who you are, but a few who I feel deserve a particular recognition. To Ian Bradbury, Adam Cook, Dan Hasselman, Megan McCusker, Friso Palstra and Craig Reynolds my many thanks for your help, lengthy discussions and the debates and laughs that accompany sharing an office and a lab.

The completion of this thesis would not have been possible without the support and time off allowed by the Rivers and Fisheries Trusts of Scotland. Particularly to Callum Sinclair and Andrew Wallace, my sincere thanks for your understanding and commitment in helping me achieve this goal.

To my family, thank you doesn't seem to say enough for your constant and unwavering emotional (and financial at times!) support. You have always believed in me, even when I perhaps did not believe in myself and have always been there for me.

Finally, I owe a true debt of gratitude to my wife, Megan. You have been on life's journey with me since before this thesis began and have put up with the laughter, despair, frustration, exhaustion, hope and joy that comes with the territory. You have taken on more than your fair share of responsibilities, especially during these last few months and did so with a positive attitude and a smile on your face - as always. Throughout it all you have both challenged and supported me and for this I will be eternally grateful.

CHAPTER 1

INTRODUCTION

1.1 HISTORICAL AND CONTEMPORARY PROCESSES SHAPING POPULATION GENETIC STRUCTURE

The ability to genetically distinguish between populations of a species can play an important role in management of biological resources as well as contribute to our understanding of a species' ecology, evolutionary past, and future potential to adapt to a changing environment. Population genetic structure represents a complex interaction of ecological, demographic, behavioural, and selective factors as well as the physical environment. Thus, a major focus of population genetic studies is to investigate how the current environment and the species' biology may promote or restrict population differentiation. Often, however, the distinction of populations is confounded by historical events, such as previous glacial cycles and subsequent recolonization patterns and processes as habitat becomes available (Hewitt 1996, 2000). The interplay between these historical and contemporary factors in shaping the genetic diversity and structure of a species can complicate the interpretation of population genetic patterns. For example, recent range expansions may result in weak population structure that could be interpreted as evidence for high levels of contemporary gene flow. On the other hand, secondary contact may demonstrate the existence of seemingly pronounced geographical genetic structuring, even in cases where introgressive hybridization between lineages occurs. For both types of situations, the time required to reach equilibrium is directly proportional to population size and inversely proportional to levels of gene flow (Whitlock 1992).

Among fishes, differences in both levels of population genetic structuring (Ward *et al*. 1994) and genetic diversity (DeWoody & Avise 2000) have been demonstrated

among freshwater, anadromous and marine fishes. The latter generally show the highest levels of diversity and lowest degree of population structure while freshwater fishes have the highest levels of structuring with the lowest level of diversity, with anadromous fishes being intermediate in character. The relative isolation of freshwater fish populations, and hence greater restriction in terms of gene flow among demes may seem obvious, however, the opposite conclusion of demographically 'open' populations of marine species has been questioned in recent years (see Hauser & Carvalho 2008 for a review). Among marine species, pelagic larval duration (PLD) allows for dispersal and gene flow among distant populations (Leis 1984), however the relationship between dispersal and gene flow has been demonstrated to also be affected by latitude, adult body size and water depth as well as egg type (Bradbury *et al.* 2008a; Lies *et al.* 2013). Furthermore, across the range of the focal species considered in this thesis, it has been shown that despite the occurrence of a larval pelagic phase, levels of genetic differentiation vary dramatically across differing spatial scales and landscapes (Bradbury *et al.* 2006, 2008b, Kovach *et al*. 2013).

While many cases of genetic population structure among marine fishes demonstrate small but significant genetic differences (Waples 1998), such differences have been shown to have environmental correlates (Bekkevold *et al.* 2005) and/or temporal stability (Cimmaruta *et al.* 2008), even occurring on small spatial scales of tens to hundreds of kilometres (Ruzzante 1998; Knutsen *et al.* 2003; Nielsen *et al.* 2004; Olsen *et al.* 2008; Hauser & Carvalho 2008). Furthermore, with the use of alternative markers, including those demonstrated or suspected to be under selection (Pogson 2001; Canino *et al.* 2005; Bradbury *et al.* 2010, Bourret *et al.* 2013), the evidence for stronger genetic structuring in

marine species has become clearer, demonstrated by an increase in the levels of genetic differentiation among populations.

One of the major factors responsible for patterns of genetic differentiation and ecological speciation observed today among fishes has been the historical legacy of previous glacial cycles (Bernatchez & Wilson 1998; Gagnon & Angers 2006). Studies of the phylogeography of many north temperate species have demonstrated varying patterns in the number and distribution of genetic lineages since recolonization from glacial refugia (e.g. Bernatchez & Dodson 1991; Wilson *et al.* 1996; Bernatchez 1997; Danzmann *et al.* 1998; McCusker *et al.* 2000; Wares & Cunningham 2001; April & Turgeon 2006; Barluenga *et al.* 2006; Maggs *et al.* 2008; Canino *et al.* 2010; Liu *et al.* 2011). Many of these studies demonstrate areas of secondary contact between two or more distinct lineages while others have shown dramatic differences comparing populations from previously glaciated versus unglaciated areas (e.g. Hasselman *et al.* 2013). The recent recolonization of previously glaciated areas supports the likely scenario that few species may have achieved a state of equilibrium (Whitlock & MacCauley 1999; Grosberg & Cunningham 2001). Assumptions of equilibrium are made by certain population genetic methods, such as the inference of number of migrants (Wright 1943, but see Whitlock & MacCaulay 1999) and the presence of isolation-by-distance (IBD) to infer the relative effects of gene flow and drift (see Hutchison & Templeton 1999), however non-equilibrium populations may also exhibit IBD patterns (Slatkin 1993).

Clearly, the consideration of both historical and contemporary influences on population genetic structure is of critical importance to our understanding of how populations diverge and are influenced by their environment. Bermingham and Avise

(1986) strongly emphasized this point by stating, “Thus, any interpretations of population genetic structure that fail to consider the possible influence of history in shaping that structure, may be seriously inadequate”.

1.2 RAINBOW SMELT

Rainbow smelt (*Osmerus mordax* Mitchill, 1815) is a small-bodied, euryhaline species occurring in waters off the coast of North America, with a contemporary native range from New Jersey (40^{o}N) north to Labrador (53^{o}N) (Scott & Scott 1988) and is an important forage species for larger piscivorous fishes and birds. The species range has expanded both via intentional and accidental introductions into the Great Lakes, Mississippi and Hudson Bay basins (Mercado-Silva *et al.* 2006), where they have contributed to the extirpation of native fishes via competitive or predatory interactions (Evans & Waring 1987, Hrabik *et al*. 1998, 2001). Within their native range, rainbow smelt are primarily considered anadromous, although numerous freshwater, landlocked populations have been identified. These latter populations have occasionally evolved into two distinct ecotypes: a large, macrophagous form and a much smaller microphagous form. The spawning runs of these two ecotypes may occur in separate or the same streams within a system the two ecotypes have evolved via independent divergences *in situ* within each system from ancestral anadromous forms (Baby *et al.* 1991; Taylor & Bentzen 1993a). The microphagous ecotype is primarily planktivorous and is characterized by a smaller size distribution (60-150 mm), more gill rakers, and relatively larger eyes compared to the macrophagous ecotype (150-300 mm). Hybridization between the forms is thought to be rare although it does occur (Saint-Laurent *et al.* 2003;

Curry *et al.* 2004; Bradbury *et al.* 2011a), and evidence has shown a decrease in fertilization success of the hybrids compared to the pure forms (Bradbury *et al*. 2010).

Among the anadromous life-history type, spawning usually takes place in early spring, with smelt entering brooks and streams at night and spawning above the head of tide. Males arrive first on the spawning grounds with larger individuals arriving earlier in the spawning run, as body size has been observed to decrease over the length of the spawning period. Smelt generally mature at 2-3 years of age and may live up to 4-5 years, spawning in each of up to three consecutive years (MacKenzie 1964). The small (~ 1 mm) eggs are adhesive and attach to gravel, vegetation and each other. Heavy egg mortality is common in dense areas of egg deposition. Incubation time varies depending upon the water temperature, with hatching occurring at 10-20 days at 14-16^{o}C. Hatching of larvae is synchronized to dark cues (Bradbury *et al.* 2004) and increased night-time abundance of drifting larvae has been observed (Johnston & Cheverie 1988). Larvae are immediately carried downstream to the estuary, with freshwater residence time of the young thus quite limited compared to most anadromous species. Population genetic structure of the anadromous form to date has shown a varied degree of genetic differentiation in different portions of the species' range (Bradbury *et al.* 2006, 2008; Kovach *et al.* 2013) and supports a member-vagrant (*sensu* Iles & Sinclair 1982) hypothesis of population structure (e.g. Bernatchez & Martin 1996). Limited movement between estuaries has also been suggested based upon otolith microchemistry (Bradbury *et al.* 2008c).

Mitochondrial DNA (mtDNA) has previously been used to identify two glacial races with a zone of secondary contact in the St. Lawrence estuary (see Bernatchez 1997) and it has been suggested that this has also led to adaptive morphological divergence among

both races and different feeding specializations (Barrette *et al.* 2009). Work presented in this thesis (Chapter 3) and described in Bradbury *et al*. (2011b) has revealed a previously unknown second zone of contact between these two races in southeast Newfoundland, a pattern also evidenced by the nuclear microsatellite data (Bradbury *et al.* 2011b).

Rainbow smelt present a dichotomy in a number of traits, allowing for investigation into the patterns and processes underlying intraspecific variation and differentiation. Their alternative anadromous *vs*. freshwater life-histories with adaptive ecological differentiation in the latter allow for investigation of the processes involved in the early stages of speciation. The presence of two glacial races allows for reconstruction of historical routes of recolonization leading to their present distribution, while contrasting levels of genetic diversity and differentiation allow for investigation into the influence played by the geography and the local environment in promoting and maintaining population differentiation. These factors along with a relatively short generation time, suggest the possibility for rainbow smelt to serve as a model species toward our understanding of ecological and evolutionary processes in shaping species diversity.

1.3 THESIS OVERVIEW

Previous genetic studies of smelt have focused primarily on the distinction between two major glacial races identified in the St. Lawrence estuary and the evolution and maintenance of the two freshwater ecotypes. The work presented here focuses on anadromous populations throughout the native range and both historical and contemporary factors influencing the genetic population structure and diversity of the species.

Chapter 2 reviews the use of several measures of summarizing genetic differentiation that have been debated recently in the literature (G_{ST}, G'_{ST}, and D). This chapter expands upon this debate with the use of simulations and conducts a meta-analysis of freshwater, anadromous and marine life-histories of fishes utilizing these different statistics for two types of genetic markers (allozymes and microsatellites). Among the different life-histories of fishes explored, the results show that traditional measures have disproportionately underestimated levels of genetic differentiation among marine fishes when using microsatellites, by as much as 10-fold. The use of the different measures is discussed and demonstrates that traditional measures may be problematic when comparing species and/or populations with different levels of genetic diversity.

Chapter 3 is a phylogeographic exploration of the native range using a portion of the mitochondrial genome previously identified in distinguishing between two glacial races of smelt. With extended geographical coverage, a second zone of secondary contact was identified in Newfoundland, in addition to the previously recognized contact zone in the St. Lawrence estuary. This chapter focuses on the distribution of the two clades and identifies intermediate haplotypes at the southern portion of the sampled range. In light of the discovery of the Newfoundland contact zone, the previous recolonization scenario is tenuous and discussion is focused on alternative explanations.

Chapter 4 reports the work of the primer development for a set of 12 di- and tetranucleotide microsatellite loci and their initial characterization for levels of diversity.

Chapter 5 uses microsatellite markers to investigate the population structure of anadromous rainbow smelt throughout much of their native range. Results demonstrate significant regional variation in both the levels of genetic diversity as well as levels of population structure. Most notable are the differences in both levels of genetic diversity

and differentiation between Newfoundland and the mainland portion of the species' range. Within the latter area, however regional differences in the scale at which genetic structuring occurs is evident. Populations within the Gulf of St. Lawrence show a very weak pattern of isolation by distance while those within the Bay of Fundy show an elevated level of genetic differentiation and a more restricted scale of dispersal. These results mirror some of the phylogeographic patterns described in Chapter 3 and further demonstrate the effect of local, contemporary forces in shaping different patterns of genetic connectivity and diversity within a species.

Chapter 6 focuses on temporal genetic differentiation or, isolation-by-time (Hendry & Day 2005) in continuous vs. discontinuous spawning runs of smelt. Two scenarios are explored with microsatellite data and the use of simulations contrasting an uninterrupted spawning run versus one with a defined break in spawning activity between an 'early' and 'late' run component. Results demonstrate a lack of temporal genetic differentiation in the continuous spawning run while the 'early' and 'late' run components show levels of genetic differences similar to that observed between spawning locations separated by >100 km.

Chapter 7 summarizes the major findings of this research and discusses the implications of the results as well as outlining areas of future work relating to smelt phylogeography and population divergence.

1.4 Publications Arising From The Thesis

At present, two chapters resulting from this thesis have been published (Chapters 4 and 6):

Coulson, M.W., Paterson, I.G., Green, A., Kepkay, R., and Bentzen, P. 2006. Characterization of di- and tetranucleotide microsatellite markers in rainbow smelt (*Osmerus mordax*). *Molecular Ecology Notes*, **6**, 942-944.

Coulson, M.W., Bradbury, I.R., and Bentzen, P. 2006. Temporal genetic differentiation: continuous *v*. discontinuous spawning runs in anadromous rainbow smelt *Osmerus mordax* (Mitchill). *Journal of Fish Biology*, **69** Supplement C, 209-216.

Additionally, results of some of the data presented in this thesis have contributed to the following publications:

Bradbury, I.R.,* Coulson, M.W.,* Campana, S.E., Paterson, I.G., and Bentzen, P. 2011. Contemporary nuclear and mitochondrial genetic clines in a north temperate estuarine fish reflect Pleistocene vicariance. *Marine Ecology Progress Series*, **438**, 207-218 (*contributed equally to this work).

Bradbury, I.R., Coulson, M.W., Campana, S.E., and Bentzen, P. 2006. Morphological and genetic differentiation in anadromous smelt (*Osmerus mordax*) along eastern Canada: disentangling the effects of distance and selection on gene flow. *Journal of Fish Biology*, **69** Supplement C, 95-114.

CHAPTER 2

SUMMARIZING POPULATION GENETIC DIFFERENTIATION AMONG FISHES: MIGRATION, MUTATION AND COMPARABILITY OF STUDIES

2.1 INTRODUCTION

Determination of the degree of allele frequency differences among demes can play an important role in the management of biological resources, conservation of threatened or declining species, as well as understanding the evolutionary and ecological interactions among populations of a species. While F_{ST} (Wright 1943, 1951) and its analogues predominate in the literature for summarizing population genetic differentiation, this parameter has two undesirable properties when divergence time and polymorphism are high: (i) it reaches an asymptote and (ii) this maximum value is inversely proportional to the amount of polymorphism present in the populations (Hedrick 1999). This limiting power of within-population diversity on F_{ST} (and its analogues) has been acknowledged for some time (Wright 1943, 1951; Charlesworth 1998; Hedrick 1999), but recently several alternative measures have been discussed (Hedrick 2005; Meirmans 2006; Jost 2008). Despite the obvious limitations of F_{ST}, it is still a desirable statistic used by most population geneticists because of its direct link, under an infinite island model, to the number of migrants exchanged between populations - a parameter of immense interest to ecologists and evolutionary biologists. Additionally, given its historical and continuing widespread use, F_{ST} remains a desirable parameter for comparative purposes (Neigel 2002). While G_{ST} (Nei 1973) is regarded as an extension of Wright's F_{ST} to the case of multiple alleles, I will refer to these two parameters interchangeably.

2.1.1 THE NATURE OF F_{ST} AND RELATED MEASURES

Theoretically, F_{ST} (and its analogues) can vary between 0 (no population differentiation) and 1 (complete differentiation, i.e. no shared alleles). One commonly used F_{ST} analogue, Nei's (1973) multilocus G_{ST} is computed as

$$G_{ST} = \frac{H_T - H_S}{H_T}, \qquad \text{(equation 1)}$$

where H_T represents the heterozygosity of the pooled populations and H_S the average heterozygosity of the subpopulations. Therefore, for the theoretical upper limit to be reached, each population could only have one allele (making $H_S = 0$) and not all populations could have the same allele (therefore $H_T > 0$), a situation unlikely for most datasets, especially highly variable markers such as microsatellites. While small populations exchanging few migrants will show relatively high levels of differentiation, larger populations with low levels of differentiation (as measured by F_{ST}) pose a potential uncertainty. It could be that these populations have low differentiation as a consequence of actually exchanging many more migrants, or that they are relatively isolated but have retained higher levels of heterozygosity. This higher level of heterozygosity could be due to a recent separation and insufficient time for genetic drift to cause differentiation, historically large effective population size, or the greater influence of mutation rate and its expectation to reduce estimates of F_{ST} (Crow & Aoki 1984; Epperson 2005).

As Jost (2008) pointed out, an F_{ST} of 0.001, for example, could indicate low, moderate, or maximal differentiation depending on the level of within subpopulation heterozygosity. Indeed, among marine fishes, for instance, it is common in the literature

to see very small (i.e. < 0.02), but yet statistically significant values of F_{ST} among populations of a species known to have very large population sizes. In this type of scenario, there has been considerable debate as to whether 'statistical significance' at such low levels of differentiation represents actual 'biological significance' (Waples 1998; Palumbi 2003; Waples & Gaggiotti 2006). However, it may be the case that these low levels of differentiation are due to the higher levels of heterozygosity in these species, and that using a measure unbiased from heterozygosity would in fact, increase the magnitude of inferred levels of differentiation. However, it has long been thought that marine species are highly connected across their ranges and therefore perhaps the low levels of genetic differentiation inferred reflect this greater exchange of individuals among demes compared to more geographically restricted species (for a review see Hauser & Carvalho 2008). Therefore, while it is important to know whether or not populations are significantly different from one another, it is at least equally important to be able to correctly *quantify* the magnitude of such differences.

Hedrick (1999) demonstrated that Nei's (1973) equation for G_{ST} could be re-written as follows:

$G_{ST} = 1 - \frac{H_S}{H_T}$, which is constrained by assuming a value < $1 - H_S$,

where $1 - H_S$ is homozygosity. He pointed out that differentiation therefore cannot exceed the level of homozygosity, no matter what evolutionary forces are influencing the amount and pattern of variation.

2.1.2 G'_{ST} AND D

Hedrick (2005) proposed a correction that calculates what the maximum G_{ST} value could be (G_{STmax}), given the observed level of heterozygosity. He reported the calculation of G_{STmax} as:

$$G_{STmax} = \frac{(k-1)(1-H_S)}{k-1+H_S}, \qquad \text{(equation 2)}$$

where k is the number of subpopulations. The observed value is then compared to this maximum value to obtain a measure (G'_{ST}) that can assume any value in the parameter range. From there, G'_{ST} is calculated as

$$G'_{ST} = \frac{G_{ST}}{G_{STmax}} = \frac{G_{ST}(k-1+H_S)}{(k-1)(1-H_S)}. \qquad \text{(equation 3)}$$

More recently, Jost (2008) developed and evaluated the use of the parameter D, a measure of differentiation that is independent of population heterozygosity. While he did not evaluate the performance of G'_{ST} directly he mentioned that qualitatively it serves a similar purpose as the D parameter, in removing effects of within population heterozygosity. Jost's D is calculated as follows:

$$D = (H_T - H_S)/(1 - H_S)[\frac{k}{k-1}]. \qquad \text{(equation 4)}$$

Jost's (2008) central tenet is that G_{ST} and related measures do not capture true differentiation and that the continued use of such statistics to this end rests, in part, on two fundamental misunderstandings about the mathematics of differentiation measures: (1) a misconception that additive partitioning of heterozygosity leads to a between-subpopulation component that measures differentiation and (2) that an increasing

measure of complexity (i.e. heterozygosity) can be equated with diversity when used in ratio similarity measures.

2.1.3 WHICH MEASURE TO USE?

Eliminating or accounting for the influence of heterozygosity on measures of genetic differentiation should make it easier to compare the amount of differentiation among different studies, loci, and organisms, comparisons that have been acknowledged as problematic in the past (Charlesworth 1998; Nagylaki 1998; Hedrick 1999). Over the last several years, however, there has been considerable attention and debate in the literature regarding the relative merits of G_{ST} vs. either G'_{ST} and/or D (Jost 2008, 2009; Ryman and Leimar 2008, 2009, Heller & Siegismund 2009; Gerlach *et al.* 2010; Leng & Zhang 2011; Meirmans & Hedrick 2011; Whitlock, 2011; Wang 2012). This dialogue has largely focused on the theoretical and mathematical expectations or simulations of the various measures, although Heller & Siegismund (2009) conducted a meta-analysis of 34 studies from the literature.

It has been argued that G_{ST} is still a valid measure, particularly in relation to linking genetic differences to demographic factors (Ryman & Leimar 2008, 2009, Whitlock 2011), as D approaches equilibrium much more slowly due to its lack of dependence on population size (Ryman & Leimar 2008; Meirmans & Hedrick 2011). Jost (2008) pointed out that the properties of G_{ST} (and its relatives) are precisely why they are, in some circumstances, useful for demographic inference, but not a good measure of overall or *true* (*sensu* Jost 2008) differentiation. The use of *true* differentiation appears to be one that describes the relative overlap or sharing of alleles among demes, similar to measures such as the allele-sharing coefficient D_{AS} (Bowcock *et al.* 1994) and therefore

is influenced by mutational as well as demographic factors. In other words, *D* can be thought of as a purely descriptive method of measuring how different the allele frequencies are among the subpopulations, which encompasses all demographic, evolutionary and marker-specific properties such as mutation (Jost 2009). Furthermore, this description does not require any assumptions to be made regarding the mutation or migration model, nor the achievement of equilibrium status. However, Whitlock (2012) recently questioned the usefulness of *D*, underlining the lack of any direct relevance to the evolutionary literature or theory and suggests that both *D* and G'_{ST} tells us little about the underlying evolutionary and demographic processes. On the other hand, he acknowledges that G_{ST} is an extremely poor measure of the evolutionary isolation of a population when mutation rate is close to or greater than migration rate, which has been acknowledged by others (Ryman & Leimar 2008 and addressed here), and cautioned against its use when using microsatellite markers. Leng and Zhang (2011) used computer simulations under a variety of scenarios and demonstrated that while initial levels of variation affected both G_{ST} and *D*, the latter appeared to be more severely affected. Furthermore, these authors showed that *D* was also more affected by the mutation model than G_{ST}.

Gerlach *et al.* (2010) demonstrated that *D* more accurately estimates known levels of differentiation than G_{ST}, and showed that G_{ST} ranks divergence between populations correctly when the number of alleles remains the same but divergence decreases (i.e. *migrational* effects). However, with increasing number of alleles (i.e. *mutational* effects), ranking based upon G_{ST} is problematic. Interestingly, and quite usefully, these authors demonstrate that comparisons of significance/non-significance between G_{ST} and *D* were almost always in agreement. Therefore conclusions based on statistical assessment of the

significance or non-significance of comparisons are unlikely to be altered by one's choice in statistic, but the absolute level of differentiation will be underestimated with G_{ST} in many situations.

2.1.4 Application of Genetic Differentiation Measures to Fishes

The aim of this study is to compare the behaviour of the three measures: G_{ST}, G'_{ST} and D using two approaches. The first approach uses a series of simulated data sets designed to contribute to the current simulations to date on this topic (e.g. Leng & Zhang 2011). While most of the discussion surrounding this debate has used (or implied) a finite island model of migration, I also include the use of a one-dimensional stepping-stone model, allowing for greater exchange of individuals among proximate locations. For these simulations I compared G_{ST}, G'_{ST} and D, focusing primarily on microsatellite markers with different levels of variability. This approach allows one to alter parameters (e.g., migration rate & mutation rate) that are rarely, if ever, known in empirical studies, to assess the effect of these parameters on all three measures. It is important, however, to keep in mind that each measure has different properties and that when differences arise between them, such properties need to be considered in relation to the question of interest. Most notable is the fact that G_{ST} as originally described is considered a *fixation* index, whereas D is concerned with measuring *differentiation.* Therefore while there may be instances when these two measures will give a similar result, differences arising between them may not necessarily represent a better or worse performance of one measure over the other, but rather provide insight into the different underlying properties of each.

The second approach uses a survey of the fish literature to compare the different measures. Fishes offer an opportunity to study the effects of these different measures in terms of comparing across studies. Their wide diversity and range of habitats contributes to some of the largest range of variability and differentiation seen across most taxonomic groups. Furthermore, while freshwater, landlocked populations are often closed to migration, marine species are expected to show a higher degree of connectivity (Waples 1998; Hauser & Carvalho 2008). However, the positive relationship between population size and genetic diversity (McCusker & Bentzen 2010) suggests that levels of genetic structuring among marine fishes may not simply be due to greater exchange of individuals among demes. The much larger population sizes of many marine species results in typically higher levels of genetic diversity and therefore a proportionally greater depression of levels of G_{ST} than can be achieved compared to freshwater fishes. Indeed population genetic studies of marine fishes have long reported weak but statistically significant differences, causing debate as to whether such small differences are biologically meaningful (e.g., Waples 1998). A survey of fishes therefore allows for a comparison of scenarios whereby migration should be absent to the other extreme of potentially well-connected populations, encompassing a wide range in levels of genetic diversity.

2.2 METHODS

2.2.1 SIMULATION EXPERIMENTS

For a variety of scenarios, I compared G_{ST} estimates with both G'_{ST} according to Hedrick (2005), and Jost's (2008) D, by simulating populations comprised of multilocus

microsatellite-like genotypes using EASYPOP (Balloux 2001). Simulations used either a finite island model or a 1-dimensional stepping-stone model with n subpopulations of constant size N, equal sex ratios and random mating. This generated diploid genotypes for 10 loci, which mutated at rate μ, according to one of three models: (1) a strict stepwise mutation model (SMM), (2) a K-alleles model (KAM), and (3) and two-phase model (TPM) with 70% of mutations following an SMM and 30% following an IAM. Each individual had a probability m of migrating to another population. Under these conditions, $N_e \approx N$ and I focus then on the terms Nm and $N\mu$ as representing both the number of migrants and mutations per generation, respectively. Simulations were initiated with minimal genetic diversity and the maximum number of alleles was set to 999 ($\mu = 10^{-2}$) or 50 (all other μ), depending on the mutation rate (see Leng & Zhang 2011), so as not to introduce homoplasy, beyond that inherent in an SMM-based model. Simulations were run for 10,000-50,000 generations to obtain equilibrium and multiple replicates per parameter set were conducted. Measures of differentiation (G_{ST}) were estimated in FSTAT (Goudet 1995), while G'_{ST} was calculated according to equation 3 above, and Jost's D was calculated according to equation 4.

2.2.2 META-ANALYSIS OF FISHES

In addition to the simulations generated, I also explored the literature to compare the different measures under investigation and assess to what degree studies may have underestimated levels of genetic differentiation. To this end, I chose to explore the effects of G_{ST}, G'_{ST} and D using freshwater, anadromous and marine fishes, as these three life-history strategies offer large differences in diversity and dispersal potential, thereby

covering a broad range of genetic differentiation values. A meta-analysis was conducted on both allozyme and microsatellite datasets. The allozyme dataset was largely compiled from the study by Ward *et al.* (1994), supplemented with data from Bradbury *et al*. (2008) and further literature searches. The microsatellite dataset was compiled from Bradbury *et al*. (2008) as well as further exploration of the literature.

Studies for inclusion were chosen based upon whether it was possible to obtain the following values: average subpopulation heterozygosity (H_S), number of populations sampled, and global G_{ST}. From these values I calculated H_T [according to Nei (1973)], D [equation 11 of Jost (2008)] and G'_{ST} [equation 5b of Hedrick (2005)]. Particularly for the microsatellite dataset, in cases where F_{ST} was calculated by authors according to Weir and Cockerham (1984) and therefore represents θ, I converted θ to G_{ST} as θ_{max} and G_{STmax} do not necessarily reach the same value (Meirmans, personal communication) and substituting values of θ into Hedrick's formula for G'_{ST} can lead to overestimates of the corrected value (data not shown). The relationship between θ and G_{ST} is $\theta = nG_{ST}/(G_{ST} + n - 1)$ (Cockerham and Weir 1987, 1993) so solving for G_{ST} allows us to calculate this statistic from published studies that report θ. Calculations of G_{ST} carried out this way using simulated datasets yield values of G_{ST} virtually identical to those calculated directly from the data (data not shown).

2.3 RESULTS

2.3.1 SIMULATION EXPERIMENTS

2.3.1.1 Migration vs. mutation

Figure 2.1 shows levels of heterozygosity for varying rates of mutation (infinite alleles model) and migration under an island model of migration. As the number of migrants (*Nm*) increased from very low rates (e.g., *Nm* = 0-0.1) to about one migrant per generation, there were substantial gains in heterozygosity. However, above the one migrant per generation threshold ['OMPG rule'; Mills and Allendorf (1996)], there was little to no gain in heterozygosity with increasing number of migrants. In contrast, for a given *Nm,* heterozygosity increased substantially with faster mutation rate, and at the highest-level (10^{-2}) heterozygosity was essentially independent of migration rate. This suggests that even for a given *Nm*, mutation rate heterogeneity among markers within a study is expected to yield very different levels of heterozygosity, which will differentially bias estimates of population differentiation by their varying depression of G_{ST} values. In addition, even when populations were completely isolated ($m = 0$) high rates of mutation (10^{-3}-10^{-2}) maintained substantial heterozygosity (>0.60) for relatively large populations ($N = 1000$). Figure 1 (b-f) shows the performance of G_{ST}, G'_{ST} and D for each of the mutation rates at each level of migration. The dotted vertical line represents the point at which $\mu = m$. To the left of this line mutation is the predominant force, while to the right, the effect of migration predominates.

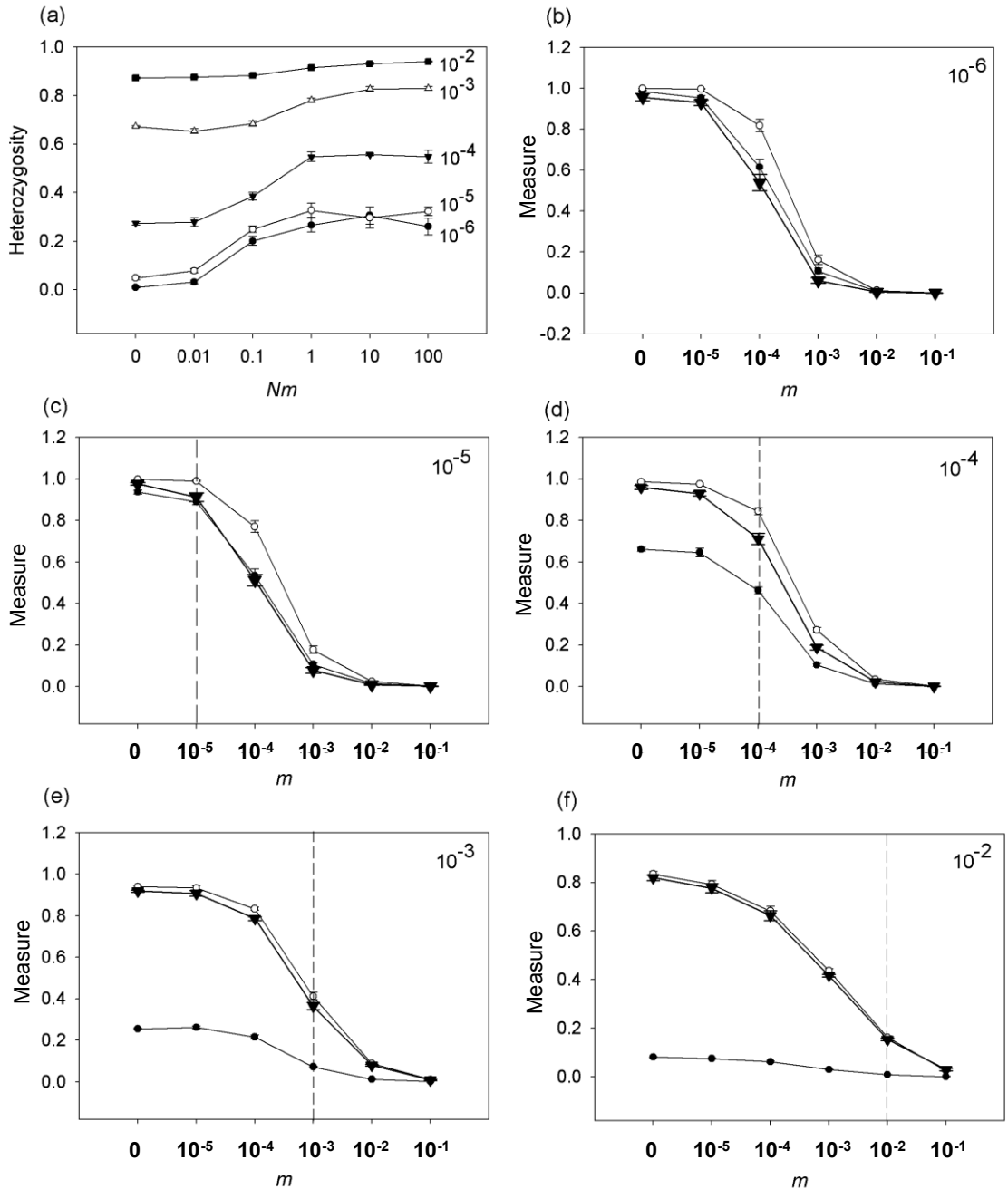

Figure 2.1 (a) Observed heterozygosity for five different mutation rates (IAM) under varying rates of migration for four subpopulations of 1000 individuals/subpopulation under an island model of migration. (b)-(f) The different measures are shown for each scenario per mutation rate. Closed circles = G_{ST}, open circles = G'_{ST}, filled triangles = D. The vertical dotted line (when present) represents the point at which $m = \mu$.

As can be seen, in the two lowest mutation rate (10^{-6} and 10^{-5}) scenarios, G'_{ST} and D produced similar estimates as G_{ST} as heterozygosity was quite low (rarely exceeded 0.30) and both G'_{ST} and D were not widely different from G_{ST}. As mutation rate increased (10^{-4} – 10^{-2}), the discrepancy between G_{ST} and G'_{ST} or D increased, and was always greatest for cases where $\mu \geq m$. It should be noted, however, that the difference could still be relatively important when the mutation rate was less than, but still a large fraction of the migration rate (i.e., just to the right of the 1:1 line).

2.3.1.2 Effect of mutation rate & model under one-dimensional isolation-by-distance

Figure 2.2 shows the trend of the three measures for four different mutation rates according to three different models of mutation, starting from a point of minimal variation and progressing toward their equilibrium values. These scenarios are shown for both a low (10^{-4}; Figure 2.2a) and high (10^{-2}; Figure 2.2b) migration rate. Under the low migration scenario, G_{ST} values for a given mutation rate was consistent under the different mutation models; however, there was considerable variation among rates within a given mutation model (range ≈ 0 – 0.8). Both G'_{ST} and D demonstrated consistent estimates for a given rate across mutation models as well, but were generally more similar across rates, with the exception of the most slowly evolving rate (10^{-5}), which took several thousand generations before accumulating mutational differences among demes. Additionally, for both these latter measures, the equilibrium values for the SMM was somewhat lower than the other two models, as would be expected due to size homoplasy involved in a stepwise model.

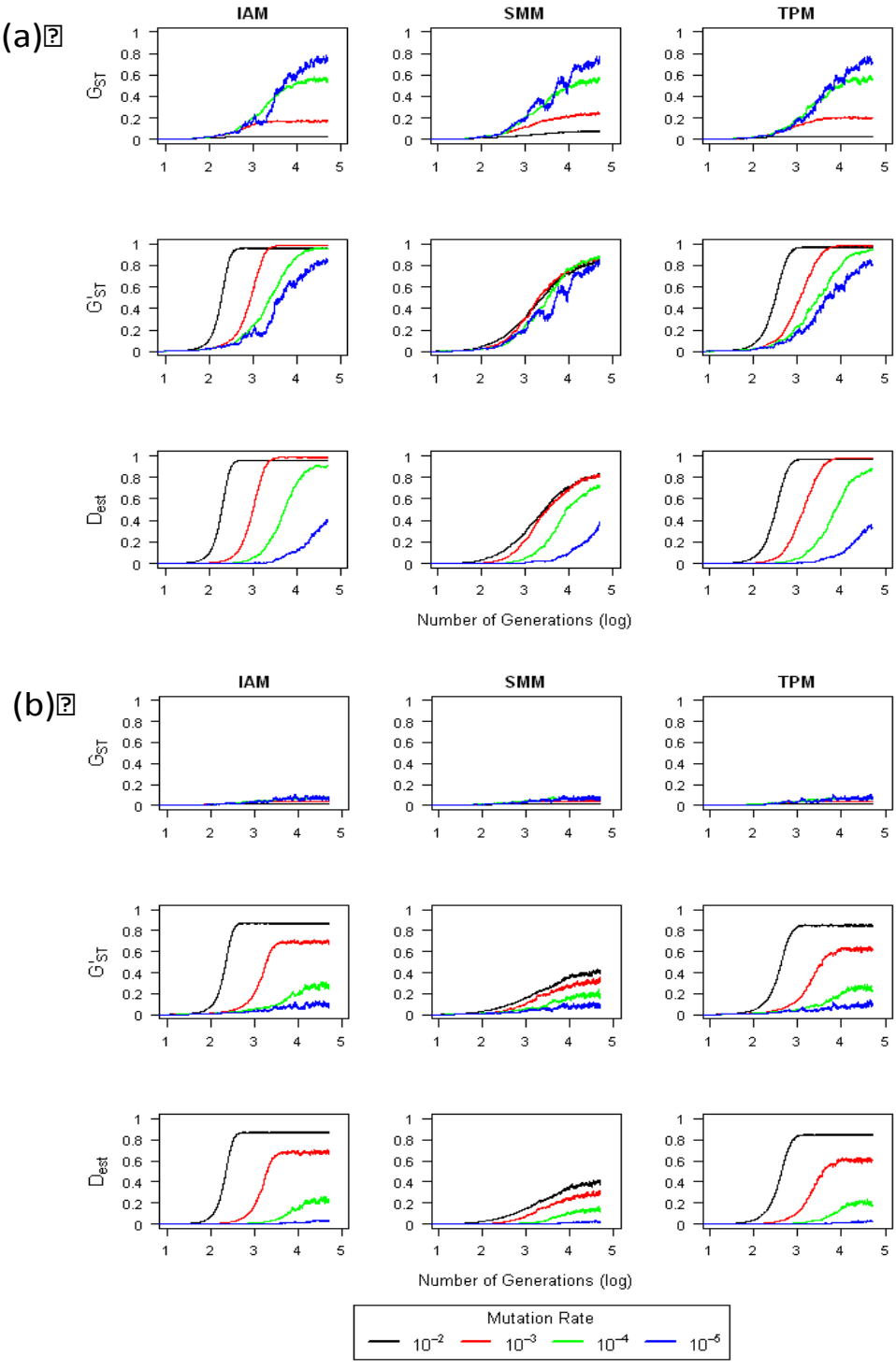

Figure 2.2 Behaviour of the three measures for varying rates of mutation under (a) a low (10^{-4}) and (b) a high (10^{-2}) migration rate for 10 populations following a 1-dimensional stepping stone model of migration.

Under a scenario of higher gene flow (Figure 2.2b), all mutation rates for G_{ST}, under each of the mutation models were consistent and relatively low (≤0.10). This time, conversely, there was greater variation among the different mutation rates for both G'_{ST} and D. The two highest mutation rates (10^{-2} & 10^{-3}) still resulted in estimates ≥ 0.6, possibly reflecting their equivalent (or nearly so) rate compared to the migration rate (10^{-2}) for this scenario. While such an exchange of migrants (Nm =10) will tend to homogenize demes, the restricted dispersal, compared to an island model will mean that any allele originating in a given population undergoes a type of serial dilution moving away from the source. Under both migration scenarios conducted here it is evident that D can take a longer time than G_{ST} (or G'_{ST}) to reach equilibrium, especially for slower mutation rates. Additionally, the initial increase in differentiation begins at different time point for either G_{ST} or D. For instance, compare the trajectories of the different mutation rates for G_{ST} and D under the low migration scenario. For the former measure, the different rates all begin to ascend toward their equilibrium values at more or less the same time point (after approximately 200-300 generations), while for D the differentiation increases at different time points depending upon the mutation rate and can occur relatively fast (≤ 100 generations) or quite slowly (~10,000 generations).

As has been noted elsewhere (Jost 2008, 2009; Ryman & Leimar 2008), G_{ST} and D differ also with respect to the relative ranking of the different mutation rates. Intuitively a measure of *differentiation* should be highest for the 10^{-2} mutation rate, as it will most rapidly accumulate differences among the populations. This however was not the case for G_{ST} values, which were inversely related to mutation rate, particularly evident under the low gene-flow scenario (Figure 2.2a). For G'_{ST} and D however,

increased levels of differentiation were observed for the higher mutation rates compared to intermediate or lower mutation rates.

Given the focus on the use of these measures with highly variable markers such as microsatellites, the following comparison used only the SMM model. For both the high and low gene flow scenarios above, I looked at the slope of the IBD across the 10 subpopulations for each of the mutation rates using the genotype data from the final generation of the simulations (i.e. at 50,000 generations) (Figure 2.3). For ease of comparison and simplicity, I focus only on G_{ST} and D. Under the higher gene flow scenario (Figure 2.3 a,b) both measures tend to show the linear increase associated with distance, as expected. However, the ranking of the mutation rates for G_{ST} is, again, generally reversed with the highest mutation rate showing a relatively flat line for G_{ST} while for D the lowest mutation rate is relatively flat. Among the remaining mutation rates, however, those for G_{ST} follow a similar slope and achieve a similar range of values while for D there is more variation among rates but with similar slope. For the lower gene flow scenario (Figure 2.3 c,d), G_{ST} tends to remain relatively flat for all mutation rate, or plateaus very early at a distance of 1 or 2 subpopulations and there is considerable variation in the magnitude of the different G_{ST} values obtained. For D under this scenario, the lower mutation rates show a linear increase while the higher rates appear to increase quickly to a plateau, with the range of values generally closer than for G_{ST}, especially at the largest geographical distances.

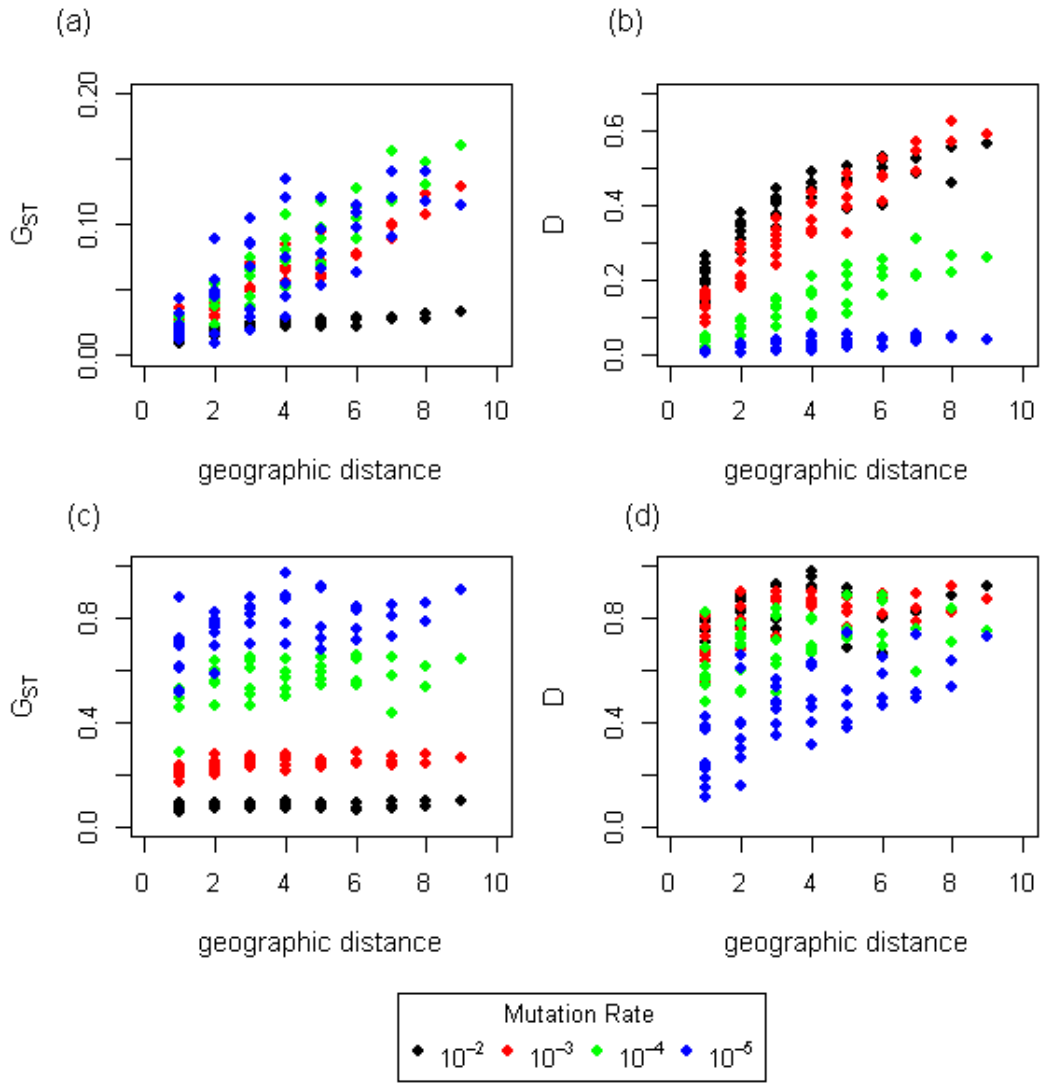

Figure 2.3 Plots of IBD using G_{ST} and D for 10 simulated populations under a one-dimensional stepping-stone model for varying rates of a stepwise mutation model under (a & b) a high gene flow model (10^{-2}) and (c & d) a low gene flow model (10^{-4}). Under the high gene flow model both measures show a linear increase however the ranking of the levels of 'differentiation' are reversed between the two measures. Under the low gene flow model, G_{ST} tends to show much more variation in both the slope and levels of 'differentiation' as compared to *D*. [Note the different scales on the *y*-axes.]

2.3.2 META-ANALYSIS OF FISHES

Comparison of the three different measures for freshwater, anadromous and marine fish data sets showed strikingly different results between the allozymes and microsatellites. As reported in Ward *et al.* (1994) for allozymes, marine fishes and freshwater fishes showed the least and greatest differentiation, respectively; whereas anadromous species showed intermediate levels of differentiation. As seen in Fig. 2.4a and 2.4c, G'_{ST} yielded virtually identical values as G_{ST} for all three life-histories and overall (average $G_{ST} = 0.138 \pm 0.189$; range = -0.002-0.852, average $G'_{ST} = 0.147 \pm 0.199$; range = -0.002-0.853) demonstrating that the latter appeared unaffected by such low heterozygosity values as is typical of allozyme data sets (average H_S across the allozyme dataset = 0.065 ± 0.051; range = 0.001-0.376). However, D (average $D = 0.017 \pm 0.035$; range = 0.000-0.318) was significantly lower than either of the former two measures (Fig. 2.4c). For example, the species *Cottus nozawae* Snyder 1911 (Andoh and Goto 1988) yielded a G_{ST} value of 0.620 and a D value of only 0.017. The relationship between G_{ST} and G'_{ST} (slope = 1.049, $r^2 = 0.997$, $P < 0.0001$) was very strong, not surprisingly, given that the latter is based on the former. However, the relationship between G_{ST} and D (slope = 0.137, $r^2 = 0.539$, $P < 0.0001$) was significant but much weaker and departed dramatically from the 1:1 relationship (Fig. 2.4a). Additionally, unlike G_{ST} and G'_{ST}, the pattern of relative levels of differentiation among life histories was not maintained with D. Freshwater fishes had the highest level of overall differentiation based upon D (0.029); however, overall levels of differentiation for both anadromous and marine fishes were the same (0.01).

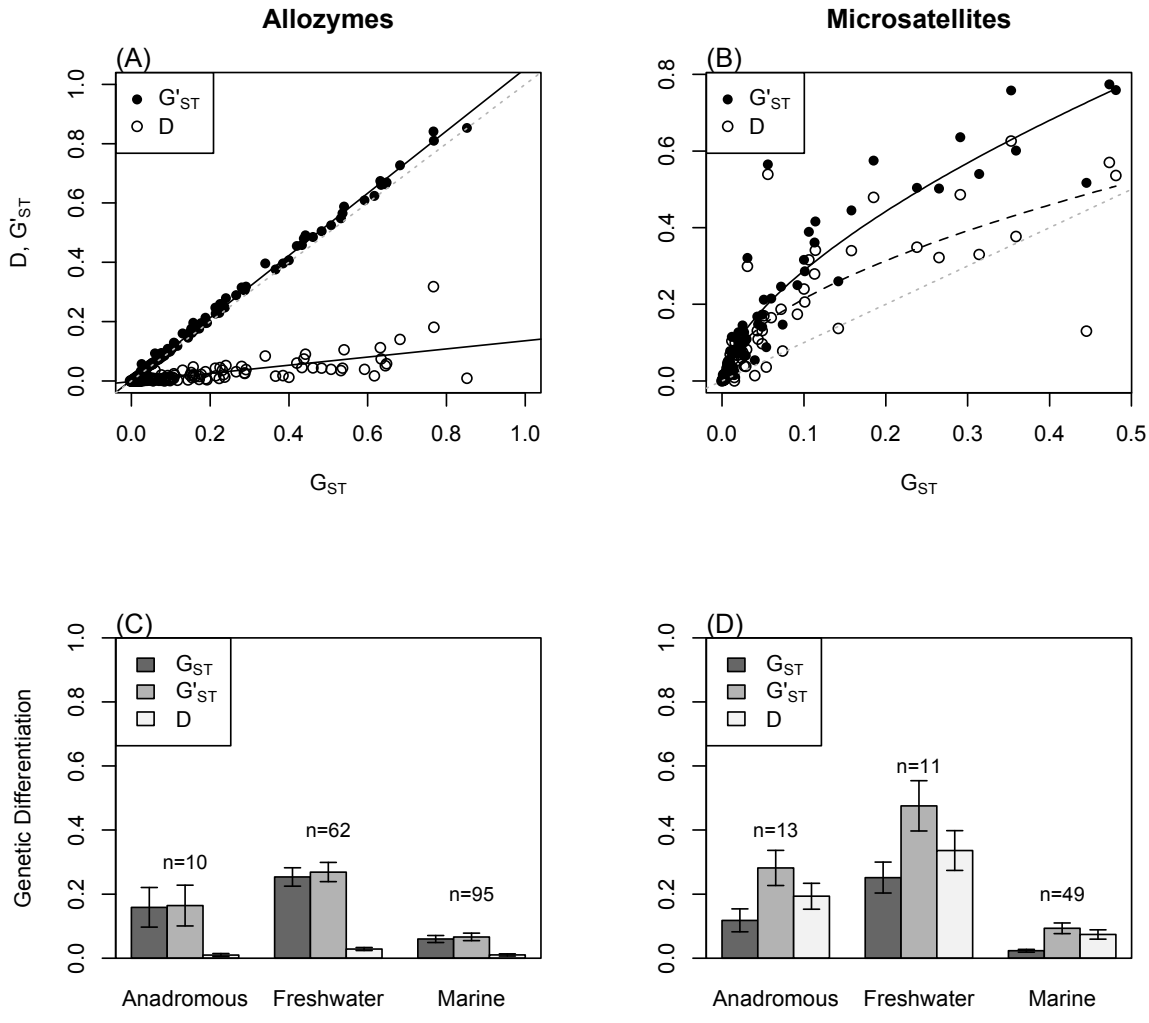

Figure 2.4 Results of a meta-analysis of genetic studies of fish species using allozymes (A) & (C) and microsatellites (B) & (D) contrasting the different measures considered. (A) and (B) show the relationships between G_{ST} and either G'_{ST} or D. In both cases, the dotted line represents the 1:1 relationship. For allozymes, a linear regression was fitted to the data, while for the microsatellites, a power function explained more of the variance. (C) and (D) show the average estimates of each measure separated by life-history type. Sample sizes for each life-history are indicated above the bars.

For the microsatellite dataset, the relative pattern of differentiation obtained with G_{ST} across life-history types mirrors that seen with allozymes (highest in fresh water, lowest in marine and intermediate among anadromous species), as also observed by DeWoody and Avise (2000) (Fig. 2.4d). For all three life history types, both G'_{ST} and D demonstrated an increase in differentiation relative to G_{ST} (Fig. 2.4b). For freshwater fishes the average G_{ST} was 0.252 ± 0.048 (range = 0.020-0.481) compared to 0.475 ± 0.079 (range = 0.088-0.774) or 0.336 ± 0.062 (range = 0.036-0.626) for G'_{ST} or D, respectively. At the other end of the spectrum, marine fishes had an average G_{ST} of 0.023 ± 0.004 (range = 0.000-0.238) compared to 0.093 ± 0.017 (range = 0.000-0.565) for G'_{ST} and 0.074 ± 0.015 (range = 0.000-0.539) for D. Anadromous species had an average G_{ST} of 0.118 ± 0.036 (range = 0.002-0.445) while estimates of G'_{ST} and D were 0.282 ± 0.055 (range = 0.004-0.575) and 0.193 ± 0.040 (range = 0.002-0.479), respectively. This translated into an average increase of about 1.3-1.8-fold for freshwater fishes, 1.6-2.4 for anadromous fishes, and 3.2-4-fold for marine fishes (Figure 2.4b). While G'_{ST} values for a given study were always equal or greater than D, the relationship between G_{ST} and G'_{ST} ($r^2 = 0.890$, $P < 0.0001$) was stronger than that between G_{ST} and D ($r^2 = 0.713$, $P < 0.0001$) (Figure 2.4d); however, this is not surprising given the dependency of G_{ST} on G'_{ST}. There was however, considerable variation in the levels of increase between G_{ST} and either G'_{ST} or D across the dataset (Figure 2.4a, b; Figure 2.5). This was most pronounced at higher levels of heterozygosity, and therefore most dramatically affected marine fishes, which in some cases showed a ~10-fold increase in levels of differentiation using either G'_{ST} or D compared to G_{ST} (Figure 2.5). While G'_{ST} was always greater than or equal to D, the difference in estimates of the two

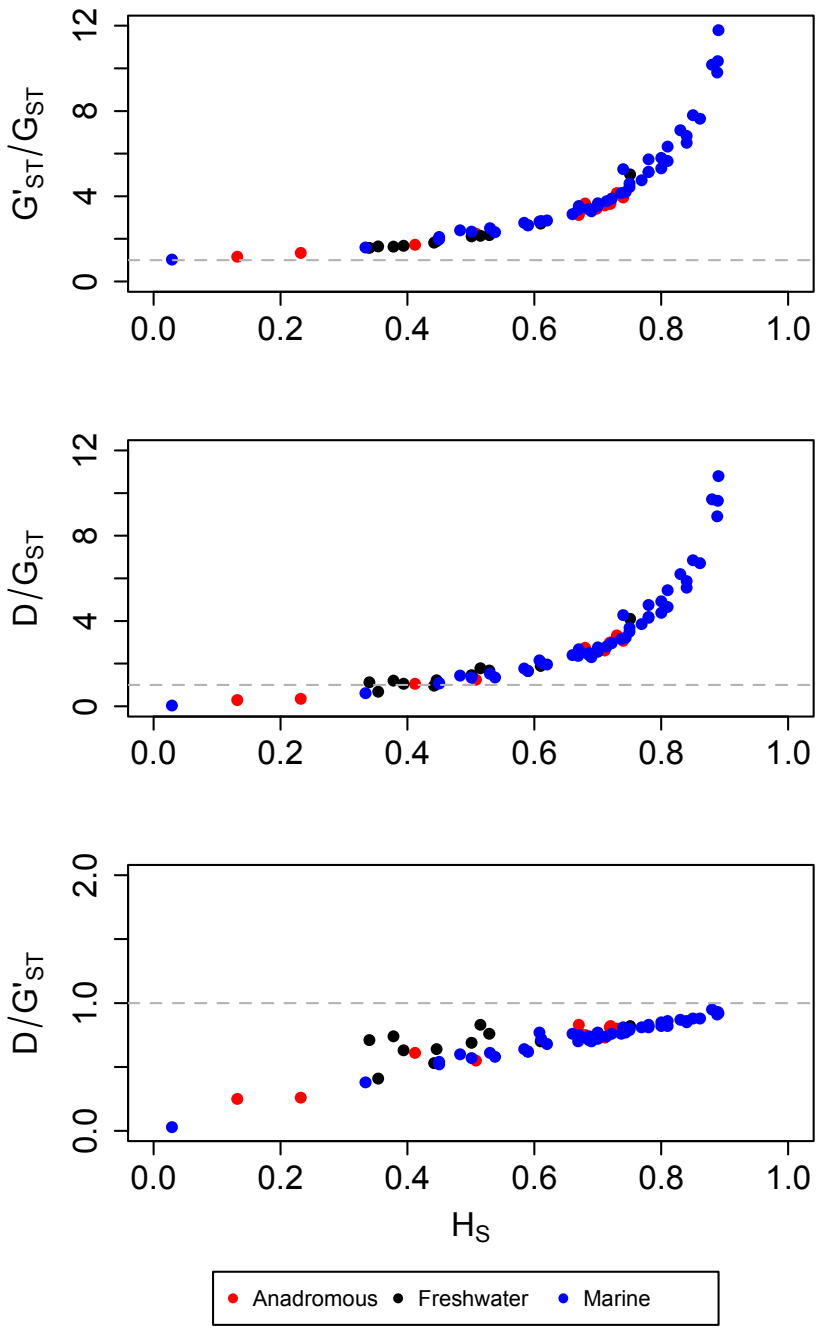

Figure 2.5 Ratios of the pairs of measures considered in the meta-analysis of fishes. In each case the three life-histories are coded by colour. The horizontal dashed line represents the 1:1 ratio.

statistics were most pronounced at lower levels of heterozygosity, while they became more similar and approached a ratio of 1:1 with increasing levels of heterozygosity (Fig. 2.5c). D was generally always higher than G_{ST}, with a few exceptions, limited to cases where at least one of the loci used was monomorphic.

Not surprisingly, there was a significant negative relationship between H_S and G_{ST} across the entire dataset (slope = -0.341, $r^2 = 0.26$, $P < 0.0001$). However, a negative relationship was still present when applying G'_{ST} (slope = -0.347, $r^2 = 0.073$, $P < 0.012$). On the other hand, D showed no relationship between its value and H_S (slope = -0.128, $r^2 = 0.007$, $P = 0.226$) demonstrating a lack of dependence on H_S (Fig. 2.6).

Given the differences in estimates of the different statistics, the question remains how well the *relative* estimates compare to one another. In other words, how well do the different measures rank these studies according to levels of differentiation, a concern illustrated by Jost (2008). To this end, Kendall's τ (Kendall 1938) was used to estimate the proportion of study pairs that are concordant in their ranking between two measures, considered for both marker types (Table 2.1). For the allozyme datasets, τ was highest, not surprisingly, for G_{ST}-G'_{ST} comparisons both overall and for each life-history type. Comparisons involving D (either to G_{ST} or G'_{ST}) were much more variable and lower, and for anadromous fishes, the ranking was not statistically significantly different from random (Table 2.1). Among the microsatellite studies the highest τ values were generally seen between G'_{ST} and D both overall and within each life-history group. Comparing either of these measures to the traditional G_{ST} yielded τ values ranging between 0.667-0.846, suggesting that approximately 15-30% of the studies would be ranked differently according to the different measures.

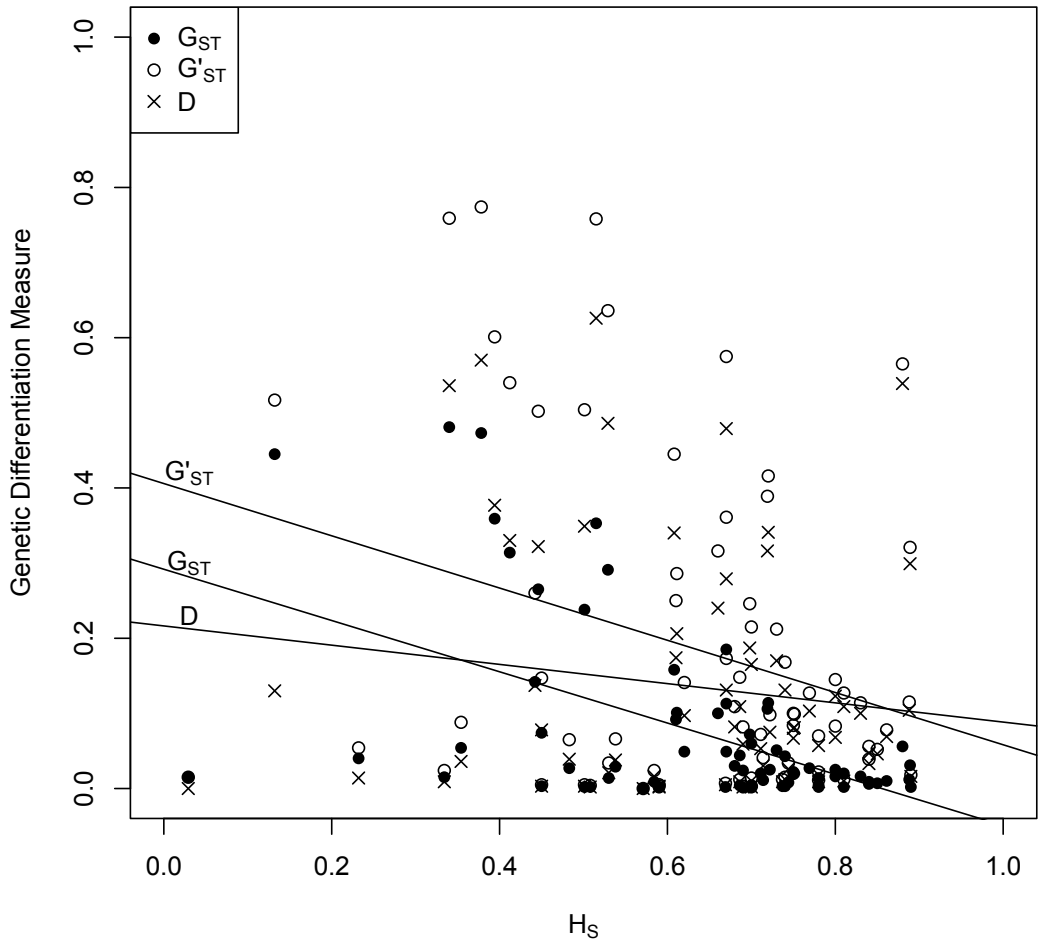

Figure 2.6 Regression of each of the three measures (G_{ST}, G'_{ST}, and D) on heterozygosity for the meta-analysis of fishes.

Table 2.1 Kendall's tau (τ) rank correlation coefficients for pairs of measures for the fish dataset overall and broken down by life-history type. All coefficients are significant at $P < 0.001$, unless otherwise noted (n.s. = non-significant).

Microsatellites	**Overall**	**Freshwater**	**Anadromous**	**Marine**
G_{ST}-G'_{ST}	0.840	0.818	0.846	0.801
G_{ST}-D	0.757	0.709	0.667	0.718
G'_{ST}-D	0.920	0.891	0.821	0.923
Allozymes				
G_{ST}-G'_{ST}	0.979	0.980	1.000	0.971
G_{ST}-D	0.722	0.672	0.477 (n.s.)	0.719
G'_{ST}-D	0.744	0.694	0.477 (n.s.)	0.749

2.4 DISCUSSION

2.4.1 COMPARISON OF MEASURES

There has been considerable debate and discussion in recent years about the relative merits and demerits of G_{ST}, G'_{ST} and Jost's D. While no single measure will be best under all possible scenarios, general guidelines have begun to emerge for the use of G_{ST} versus D (see Meirmans & Hedrick 2011 for an in-depth review). Work presented here and elsewhere has highlighted differences between these measures under different mutation models, migration rates and demographic scenarios.

Previous studies have acknowledged that alternative measures are likely to be most beneficial when mutation exceeds migration, in other words, when $4N_e\mu \geq 4N_em$ (Figure 2.1) as has been described by others (e.g., Hedrick 2005; Ryman and Leimar 2008). Furthermore, when migration greatly exceeds mutation, the different measures generally give consistent estimates, as dispersal in these situations will homogenize alleles at faster rate than they are introduced via mutation. Of course, one rarely knows *a priori* the relative impact of these forces. Moreover, even when μ was slightly less than m, there was still a difference that could lead to different conclusions about the magnitude of differentiation (Figure 2.1). Application of a threshold value for heterozygosity at which the use of either of the alternative measures (G'_{ST} or D) becomes important may be arbitrary, but from the meta-analysis of fishes (Figure 2.5) it appears that above $H_S \approx 0.6$, G_{ST} may yield significantly lower estimates when interpreted as a measure of *differentiation*.

G'_{ST} and D appear to be more affected by the particular mutation model (Figure 2.2) and this behaviour for D has also been acknowledged by others (Leng & Zhang

2011). Furthermore *D* has been criticized for being more sensitive to initial levels of heterozygosity (Ryman & Leimar 2008) and its lack of dependence on population size (Whitlock 2011), reflected in the longer time for *D* to reach equilibrium. Conversely, the negative association of G_{ST} with heterozygosity has underscored the criticism that it does not represent a measure of *true differentiation* (Jost 2008, 2009).

2.4.2 FIXATION *VERSUS* DIFFERENTIATION

The historical popularity of G_{ST} (and its relatives) is at least partly due to its theoretical link to the absolute number of migrants (*Nm*), a parameter allowing for demographic inference underlying observed subpopulation differentiation. On the other hand, *D* depends on the ratio of migration to mutation and is largely independent of population size (Jost 2008; Meirmans & Hedrick 2011). However, Wright's (1951) equation linking G_{ST} to *Nm* does include a term for mutation that has generally been considered negligible, thereby effectively leaving only the link to *Nm*. Furthermore, the assumptions underlying such a relationship have been criticized for some time (Whitlock & McCauley 1999).

The recent paper by Whitlock (2011) was entitled "G'_{ST} and D do not replace F_{ST}", and this title is entirely appropriate when considering the underlying assumptions and development of the different measures. As mentioned earlier, G_{ST} (and its relatives) was developed as a *fixation* index and therefore can be considered as a measure of tendency towards monomorphism (Gregorius 2010). Jost, on the other hand, developed *D* as a *true* measure of differentiation, to reflect both demographic and mutational effects. The discrepancy between these measures, when both are viewed as measures of *differentiation* at first may appear problematic. However, when they are viewed as

measures of *fixation* and *differentiation*, respectively, they both behave as intended. For instance, the negative association of G_{ST} and heterozygosity when viewed as a measure of *fixation* is intuitive. Lower mutation rates will not continue to replenish the genetic variation as fast as higher mutation rates will, and therefore markers with low mutation rates will be more likely to tend towards higher values. Conversely, a measure of *differentiation* should tend to be low when mutation rates are low, and take time to accumulate appreciable levels of diversity before increasing, assuming that population size is large enough to reduce the effects of drift. This is evident in the simulation results for the different mutation rates (Figure 2.4) as well as the meta-analysis, particularly in the allozyme dataset. The large discrepancy between G_{ST} and D for allozymes is explained when interpreted as *fixation* and *differentiation.* Given the low heterozygosities for typical allozyme datasets considered here, most populations would be dominated by a single allele and therefore should have a large *fixation* measure. However, a measure of *differentiation* would be expected to be quite low (as seen with D), as most individuals would have the same allele with very few alleles in the entire population.

It has been suggested that although applying traditional F_{ST} analogues to markers such as microsatellites with high heterozygosity may be problematic, the application of such measures to lower variability markers (e.g. SNPs) is likely to avoid such biases. However, the same distinction between *fixation* and *differentiation* applies. This is perhaps best emphasized with a simple hypothetical example. Figure 2.7 represents a series of examples of 10 populations under an island model spanning the range of scenarios from every population being fixed for the same allele followed by nine

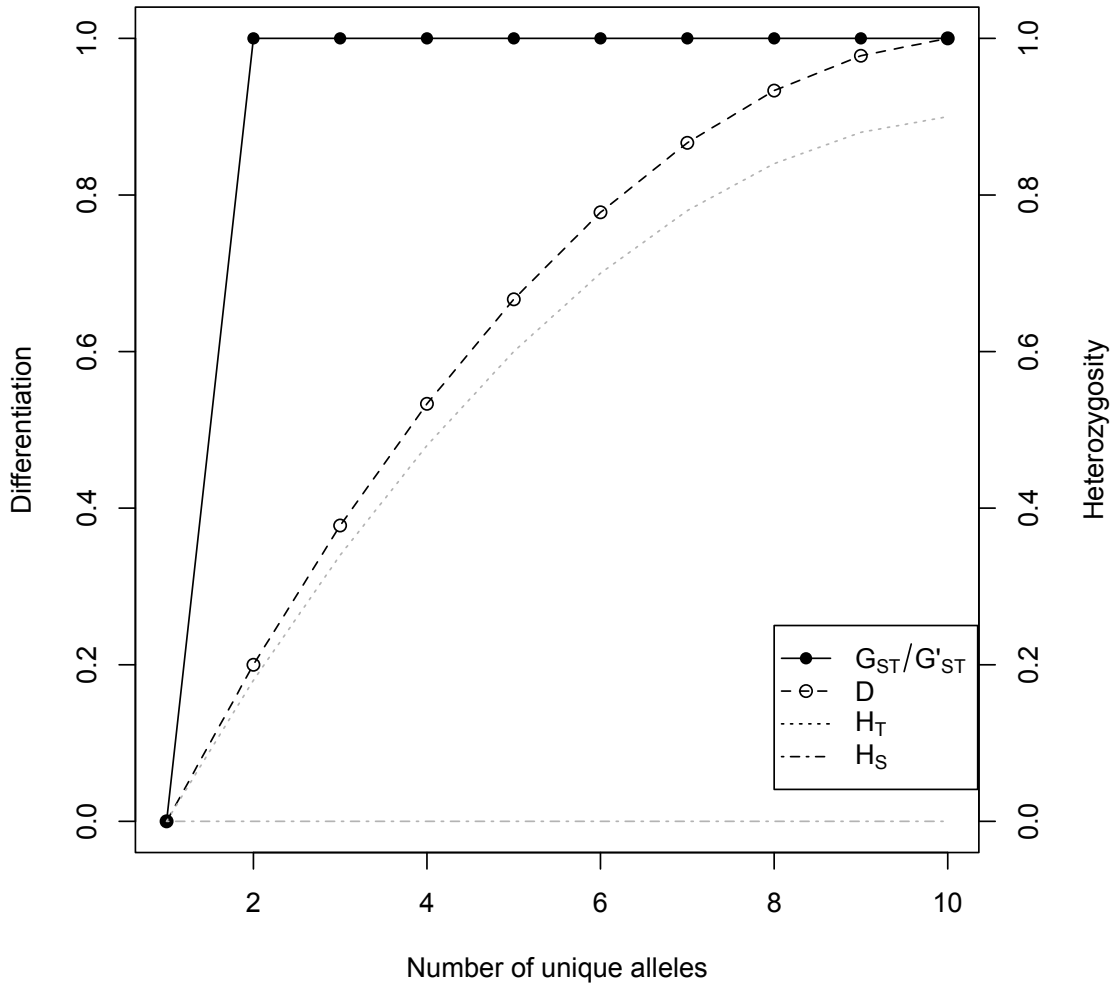

Figure 2.7 Comparison of the three measures under different scenarios with different numbers of unique alleles in the total population. With one unique allele (i.e. all populations fixed for the same allele), all measure are zero while at 10 unique alleles, each subpopulation is fixed for a different allele. Note that the lines for G_{ST} and G'_{ST} are identical. In all scenarios, $H_S = 0$ and H_T increases with increasing number of alleles in the global population.

populations sharing the same allele and the 10th population with a different, unique allele. This is repeated until all populations are fixed for unique alleles. All three measures are initially zero and *D* increases monotonically with increasing unique populations, whereas G_{ST} and G'_{ST} are both at 1.0 for all remaining scenarios (i.e. 2 – 10 unique alleles). Interpreted as a measure of *differentiation*, both G_{ST} and G'_{ST} are misleading as they would imply all scenarios are equally differentiated; however, as measures of *fixation*, they correctly reflect that in each of scenario, all populations are fixed, and assuming that one can ignore mutation in this case, they would correctly infer complete demographic isolation among the subpopulations. This highlights the fact that G'_{ST} does not represent a corrected measure of *differentiation* but rather a corrected measure of *fixation*.

Ryman and Leimar (2008, 2009) have argued that G_{ST} still has a use, and Jost (2008, 2009) pointed out that under certain circumstances, G_{ST} is a good estimate of migration and that it is precisely this property that makes G_{ST}, however, a poor measure of true differentiation as it typically ignores the effect of mutation. While this discussion has been central to the debate (see Whitlock 2011), I would argue that there is perhaps too much focus on this distinction. Ignoring the various assumptions that go along with linking a genetic summary statistic with gene flow, there are more sophisticated ways to begin to understand the underlying demographic factors influencing population genetic structure. With the advances in analytical techniques and the Bayesian revolution (Beaumont & Rannala 2004), relying on a single average statistic to infer underlying processes is less common in favour of approaches utilizing the entire data set. For instance, while it is possible to use the slope of the IBD between F_{ST} and geographic distance to infer dispersal distance (Rousset 1997; Bradbury & Bentzen 2007), other methods such as spatial autocorrelation make more use of the individual associations of

allele frequencies across the landscape. Furthermore assignment tests, using complete multilocus genotypes can be used to infer restricted dispersal. In one example, Castric & Bernatchez (2004) found significant evidence for reduced spatial dispersal associated with geographic distance in two salmonids based on genetic assignment, whereas there was no such evidence with the use of F_{ST} regressed on geographic distance. It appears that G_{ST} is better suited for inferring migration than D, however the latter is more appropriate for estimating *true* differentiation. However, the point remains as to which of these, migration or differentiation, is of most importance to the investigator. For instance, with respect to IBD patterns, is it a case of inferring dispersal distance or measuring how differentiation changes across the landscape (Meirmans & Hedrick 2011)?

2.4.3 META-ANALYSIS OF FISHES

The meta-analysis presented here demonstrated multiple cases where the use of either G'_{ST} or D reflected increased estimates of genetic differentiation. This would suggest that there are indeed natural systems where the effect of mutation rivals or exceeds that of migration. The use of either G'_{ST} or D did not change the overall interpretation regarding the degree of structuring among the different life histories of fishes. However, given that marine fishes showed the largest proportional increase for the microsatellite dataset, this suggests, on average, that the influence of higher levels of variation is higher in marine fishes than either anadromous or freshwater and that realizing 'true differentiation' may be most important for this group. Given the small but significant F_{ST}s typically associated with marine fishes (e.g. < 0.02) and the uncertain biological importance surrounding such values (Hauser and Carvalho 2008; Waples 1998; Waples and Gaggiotti 2006), accurate quantification of differentiation in these

situations will likely be most meaningful as they may assume higher values of differentiation reflective of moderate to high degrees of population structure.

The lower correlation between G_{ST} and G'_{ST} seen for microsatellite data compared to that observed for allozymes was striking. The differences observed in variance may be driven by differential sensitivity to situations of high heterozygosity, depending on the relative magnitude of mutation vs. migration. The relationship was less variable for allozymes because heterozygosity is effectively negligible and at that point $G'_{ST} \approx G_{ST}$. For situations where mutation is negligible, G_{ST} effectively does measure differentiation, and is probably one of the reasons that use of this parameter became entrenched for this purpose. A similar pattern was seen with D whereby there was a very weak, albeit significant, slope for the allozyme dataset while a much stronger relationship existed for microsatellites, but with much more scattering of the data points.

A recent meta-analysis by Heller and Siegismund (2009) across 34 studies from a variety of taxa published in *Molecular Ecology* found qualitatively similar results as the meta-analysis here. In their study they initially found both alternative measures to be negatively dependent on heterozygosity; however, when three studies that showed low within-population diversity combined with very high differentiation were removed, the negative trend between H_S and the two alternative measures disappeared. It should be noted that prior to removal of the three studies, G_{ST} showed the greatest correlation with H_S, followed by G'_{ST} and finally, with D having shown the least association. The same dependence of both G_{ST} and G'_{ST} on H_S and a lack of dependence for D was evident for the fish dataset. Also, similarly, between the two meta-analyses was the degree of incorrect rankings of the various datasets (at least for microsatellites as only they were considered by Heller & Siegismund), suggesting that when one is interested in at least

comparing the relative amounts of genetic differentiation among studies or species the choice of which measure to use may be important.

2.4.4 CONCLUSION

The emerging consensus is that traditional F_{ST} based approaches to quantifying *differentiation* may be seriously flawed and that alternative measures are needed to avoid the well-known limitations of this popular approach. As many different measures of population differentiation have been proposed and compared in the past (e.g., Goldstein *et al.* 1995; Kalinowski 2002; Paetkau *et al.* 1997; Ruzzante 1998), almost inevitably the conclusion is that no one metric will perform superiorly under all circumstances and it is therefore incumbent upon the researcher to choose based upon the dataset and particular question of interest. While the link between G_{ST} and demographic factors such as migration and population size is attractive, it also depends upon a number of assumptions unlikely to apply in most situations. Therefore the growing number of alternative analytical approaches that make use the entire dataset, rather than summary statistics, should be preferred. At a minimum, the comparison of these traditional approaches with a measure of *true* differentiation may shed light into the relative roles of demographic vs. mutational (or locus-specific) effects, and could serve as a benchmark before making demographic inferences relating to population genetic structure. While the reporting of traditional F_{ST} may still be desirable for historical comparisons, ultimately the appropriate use of alternative measures will offer new insights and highlight patterns of differentiation that may otherwise have gone unnoticed.

CHAPTER 3

PHYLOGEOGRAPHY AND RE-COLONIZATION OF TWO RACES OF RAINBOW SMELT (*Osmerus mordax*) WITH EVIDENCE FOR TWO ZONES OF SECONDARY CONTACT

3.1 INTRODUCTION

The field of phylogeography focuses on historical aspects of present-day spatial distributions of individual gene lineages (Avise 1998), and patterns are therefore influenced by both the physical characteristics of the environment and the biological and life-history characteristics of the species. For many species, emphasis has been placed on the role of previous glaciation events in structuring present-day distributions and evolutionary trajectories and indeed historical processes are increasingly implicated when interpreting contemporary population structure and connectivity (e.g. Duvernell *et al*. 2006, Orsini *et al*. 2008; McCusker & Bentzen 2010; Bradbury *et al*. 2011). The Pleistocene glaciations are believed to be the most significant historical events to have occurred during the evolutionary lifespan of most extant species (Bernatchez & Wilson 1998). The Pleistocene covers the time period between approximately 1.8 million years ago (MYA) until about 10,000 years before present (YBP). During this time approximately 20 glaciation events occurred, each lasting about 100,000 years with interglacial periods of 10,000-12,000 years (Hewitt, 2000). The most recent glaciation event, the Wisconsinan, reached maximal glaciation about 18,000 years ago with deglaciation commencing approximately 12,000-15,000 years ago (CLIMAP 1976). The advance of the glacial sheets caused widespread habitat destruction, displacement, and extirpation of location populations (Pielou 1991), and species often persisted in fringe habitats along glacial margins. Indeed, for many north temperate species, a number of

glacial refugia have been identified and implicated in species' persistence and as sources of recolonization upon glacial retreat. Therefore, intraspecific phylogeographic patterns are thought to correspond to the location(s) of recognized glacial refugia (Avise *et al.* 1987; McPhail & Lindsey 1970; Crossman & McAllister 1986) and allow for the possibility of tracing routes of recolonization (Bernatchez & Wilson 1998; Taberlet *et al.* 1998). Glacial episodes affected climate and sea level, and watercourses were altered by headwater capture and exchange (Strange & Burr 1997), factors that influenced both glaciated and non-glaciated areas. South of the ice sheets, despite individual species patterns, concordant phylogeographical signals have been documented (Bermingham & Avise 1986; Soltis *et al.* 2006) in certain geographical regions.

In the aquatic realm, it might reasonably be expected that marine species exhibit weaker phylogenetic signals, especially for highly mobile species (Avise 2000). In contrast, the freshwater environment has been more subject to habitat discontinuities, thereby increasing the potential for population structuring for freshwater-dependent species (Pielou 1991; Bernatchez & Wilson 1998). However, marine and coastal, intertidal species have indeed shown significant levels of phylogeographic structuring against a backdrop of more continuous habitat on both large, trans-oceanic scales as well as isolation in northern periglacial refugia (e.g., Wares & Cunningham 2001; Maggs *et al.* 2008). As a result of these repeated cycles, levels of genetic diversity in species diversities in formerly glaciated areas are generally below those observed in neighbouring non-glaciated regions; however, species in the former regions may have much larger ranges due to glacial melt water and the formation of large proglacial lakes that facilitated dispersal across broad scales (McAllister *et al.* 1986; Gagnon & Angers 2005). The repeated glacial episodes experienced by northern species likely contributed

to the loss of the majority of intraspecific genetic diversity resulting often in both lower diversity and differentiation compared to species or populations inhabiting non-glaciated areas (Avise *et al.* 1984; Bernatchez & Wilson 1998; Wilson & Hebert 1998). The occurrence of multiple glacial episodes may have promoted long-term diversification if the partitioning into separate refugia has been retained across multiple glacial episodes (Hewitt 1996), or alternatively the remixing of previously isolated genetic lineages may wipe out such divergences (Coope 1979; Dodson *et al.* 2007).

Due to the rapid evolution, maternal inheritance and lack of recombination, mitochondrial DNA (mtDNA) has historically been the marker of choice for phylogeographic studies (Avise 1998). It has however been acknowledged that these characteristics may also pose limitations or biases for inferences made about demographic and the evolutionary history of species (Ballard & Whitlock 2004). Increasingly, therefore, phylogeographic understanding of a species distribution and relationships is taking advantage of combining both mtDNA and nuclear genetic markers (e.g. Barluenga *et al.* 2006; Bos *et al.* 2008; Aboim *et al.* 2010; Bradbury *et al.* 2011).

In a number of cases, zones of secondary contact have been identified among distinct evolutionary lineages recolonizing from different glacial refugia (e.g. Bernatchez & Dodson 1991; Wilson & Hebert 1998; Turgeon & Bernatchez 2001). In such cases, this secondary contact may reveal cryptic levels of intraspecific differentiation or may coincide with obvious morphological differences and/or sub-species (April & Turgeon 2006). Combining different molecular markers with differing rates of evolution may further help to shed light on the status of introgressive hybridization in such contact zones. Studies on North American fishes have demonstrated that zones of secondary contact may vary from non-existent or minimal (e.g., Bernatchez & Dodson 1991, 1994;

Wilson *et al.* 1996) to geographically widespread, and can involve multiple glacial races (Wilson & Hebert 1996; Turgeon & Bernatchez 2001).

Rainbow smelt (*Osmerus mordax*) have previously been recognized as comprising two glacial races (Baby *et al.* 1991; Taylor & Bentzen 1993; Bernatchez 1997), termed Acadian ('A') and Atlantic ('B') (Bernatchez 1997), based on restriction length fragment polymorphism (RFLP) analysis on mtDNA. The distribution of these two races is independent of life-history type as both anadromous and lacustrine forms have been found to be composed of both 'A' and 'B' glacial races; however, Barrette *et al.* (2009) demonstrated suites of morphological traits associated with either vicariant history or contemporary feeding specializations. The proposed refugial origins for the Acadian race was the exundated Grand Banks off the coast of Newfoundland, as this race was initially found to predominate throughout the east coast of Canada and the Maritime provinces. The Atlantic race was predominant along the north shore of the St. Lawrence River and the only race present in the Hudson River Valley and inland regions to the north. Based on this spatial pattern, it has been postulated that the Atlantic race was associated with an Atlantic coastal plains refugium east of the Appalachian Mountains (Bernatchez 1997). The Mississippian refugium is not considered to have been a likely refugium for rainbow smelt, due to the fact that the species does not naturally occur in the Great Lakes region. A zone of secondary contact between the two races has been documented occurring along the north vs. south shores of the St. Lawrence River estuary, in which the 'A' race predominates on the south shore, and 'B' race on the north shore, despite the absence of a physical barrier to dispersal (Bernatchez 1997). Data presented in this chapter and presented in Bradbury *et al.* 2011 demonstrated the widespread existence of the Atlantic ('B') clade in southeast Newfoundland, along the Avalon

Peninsula. Clearly the occurrence of a second zone of secondary contact raises the need to revisit the current recolonization scenario of rainbow smelt since the last glaciation in light of this new evidence.

This chapter extends upon the previous RFLP phylogeographic studies of rainbow smelt by conducting direct sequence analysis, in combination with a simplified RFLP screening of an informative portion of the mitochondrial genome, the NADH5 gene. Furthermore, the current survey extends the geographical coverage of the range of rainbow smelt sampled, nominally in the province of Newfoundland and Labrador (as reported additionally in Bradbury *et al.* 2011), for which only two locations have previously been sampled. The focus of the current analysis is to further clarify the timing of differentiation of the clades and shed further light on the routes and patterns of recolonization.

3.2 MATERIALS AND METHODS

3.2.1 STUDY AREA

Overall, 50 locations in northeastern North America were sampled during 2004-2009 for consideration in the phylogeographic analysis of rainbow smelt (Table 3.1; Figure 3.1). A subset of sites was chosen for which a small number of individuals were selected for direct NADH5 sequencing (Table 3.1). Much of the focus was on further sampling within Newfoundland compared to previous studies. Additional samples from both these and other sites were screened for an RFLP, diagnostic for the major haplogroups identified (see below). In addition, data from earlier work (e.g. Baby *et al*

Table 3.1 Sites screened in the current study for both direct DNA sequencing and RFLP analysis with mtDNA group membership.

		Sequence data (723 bp)							RFLP			Total			Proportion		
Location	Sample Code	N	Number of haplotypes	*h*	π	'A' clade	'B' clade	other	N	A clade	B clade	A	B	other	A	B	other
Fore River	1	19	13	0.953	0.0047	16	1	2				16	1	2	0.84	0.05	0.11
Lake Champlain	2	10	2	0.200	0.0003	0	10					0	10		0.00	1.00	
Richelieu River	3	13	4	0.526	0.0015	1	12					1	12		0.08	0.92	
Lac St Pierre	4	8	3	0.679	0.0023	1	4	3				1	4	3	0.12	0.50	0.38
Riviere aux Outardes	5	8	4	0.643	0.0031	6	2					6	2		0.75	0.25	
Riviere Fouquette	6	6	2	0.333	0.0018	5	1					5	1		0.83	0.17	
Kennebecasis	7								24	24	0	24	0		1.00	0.00	
Maccan	8								24	22	2	22	2		0.92	0.08	
Portapique	9	8	5	0.857	0.0034	6	2		16	12	4	18	6		0.75	0.25	
Kennetcook	10								15	15	0	15	0		1.00	0.00	
Gaspereau	11								24	23	1	23	1		0.96	0.04	
Young's Brook	12								24	23	1	23	1		0.96	0.04	
Tusket River	13								22	22	0	22	0		1.00	0.00	
IndFoxPt	14	8	2	0.250	0.0004	8	0		16	16	0	24	0		1.00	0.00	
Porter's Lake	15								20	20	0	20	0		1.00	0.00	
Mosher's Brook 'early'	16a								23	23	0	23	0		1.00	0.00	
Mosher's Brook 'late'	16b								24	21	3	21	3		0.88	0.13	
Steven's Brook 'early'	17a								23	23	0	23	0		1.00	0.00	
Steven's Brook 'late'	17b								22	20	2	20	2		0.91	0.09	
Jeddore	18								24	24	0	24	0		1.00	0.00	
Eskasoni	19	7	4	0.857	0.0034	5	2		17	14	3	19	5		0.79	0.21	
Mira River	20								24	21	3	21	3		0.88	0.13	
North Harbour	21	8	5	0.786	0.0028	7	1		16	15	1	22	2		0.92	0.08	

Table 3.1 *continued*

Knoydart	22								22	22	0	22	0		1.00	0.00	
Lochaber	23	12	2	0.167	0.0002	12	0					12	0		1.00	0.00	
Wright's Creek	24								23	20	3	20	3		0.87	0.13	
Piper's Creek	25								19	11	8	11	8		0.58	0.42	
Granville	26								23	17	6	17	6		0.74	0.26	
Tyne Valley	27	8	6	0.893	0.0047	5	3		16	13	3	18	6		0.75	0.25	
Restigouche	28								24	19	5	19	5		0.79	0.21	
Magdalen Is.	29	6	1	0.000	0.0000	6	0		17	13	4	19	4		0.83	0.17	
Pt Amal	30	11	6	0.727	0.0026	10	1		7	7	0	17	1		0.94	0.06	
Conne	31	4	1	0.000	0.0000	4	0		20	18	2	22	2		0.92	0.08	
Little River	32								24	18	6	18	6		0.75	0.25	
Garnish	33	1	1	N/A	N/A	1	0		20	20	0	21	0		1.00	0.00	
Salt Pond	34	3	1	0.000	0.0000	3	0		18	17	1	20	1		0.95	0.05	
North Harbour Placentia Bay	35								22	0	22	0	22		0.00	1.00	
Long Harbour Placentia Bay	36								23	0	23	0	23		0.00	1.00	
SEPL	37	13	2	0.385	0.0005	0	13		10	0	10	0	23		0.00	1.00	
North Harbour St. Mary's Bay	38	10	2	0.200	0.0003	0	10		14	0	14	0	24		0.00	1.00	
Colinet	39	9	3	0.417	0.0006	0	9		13	0	13	0	22		0.00	1.00	
Salmonier	40	17	2	0.221	0.0015	2	15		7	0	7	2	22		0.08	0.92	
Holyrood	41	16	1	0.000	0.0000	0	16					0	16		0.00	1.00	
Biscay	42	13	1	0.000	0.0000	0	13		11	0	11	0	24		0.00	1.00	
Gull Pond	43	14	1	0.000	0.0000	14	0					14	0		1.00	0.00	
Backriver Pond	44	12	3	0.530	0.0008	0	12					0	12		0.00	1.00	
Conaughtman's Pond	45	6	1	0.000	0.0000	0	6					0	6		0.00	1.00	
Salmon Cove	46	19	4	0.380	0.0012	1	18		6	1	5	2	23		0.08	0.92	
Traverse	47	6	2	0.333	0.0023	1	5		7	0	7	1	12		0.08	0.92	
Gambo	48								22	0	22	0	22		0.00	1.00	
St. Anthony	49	3	2	0.667	0.0037	2	1		18	16	2	18	3		0.86	0.14	
Labrador	50	12	2	0.167	0.0002	12	0		12	12	0	24	0		1.00	0.00	

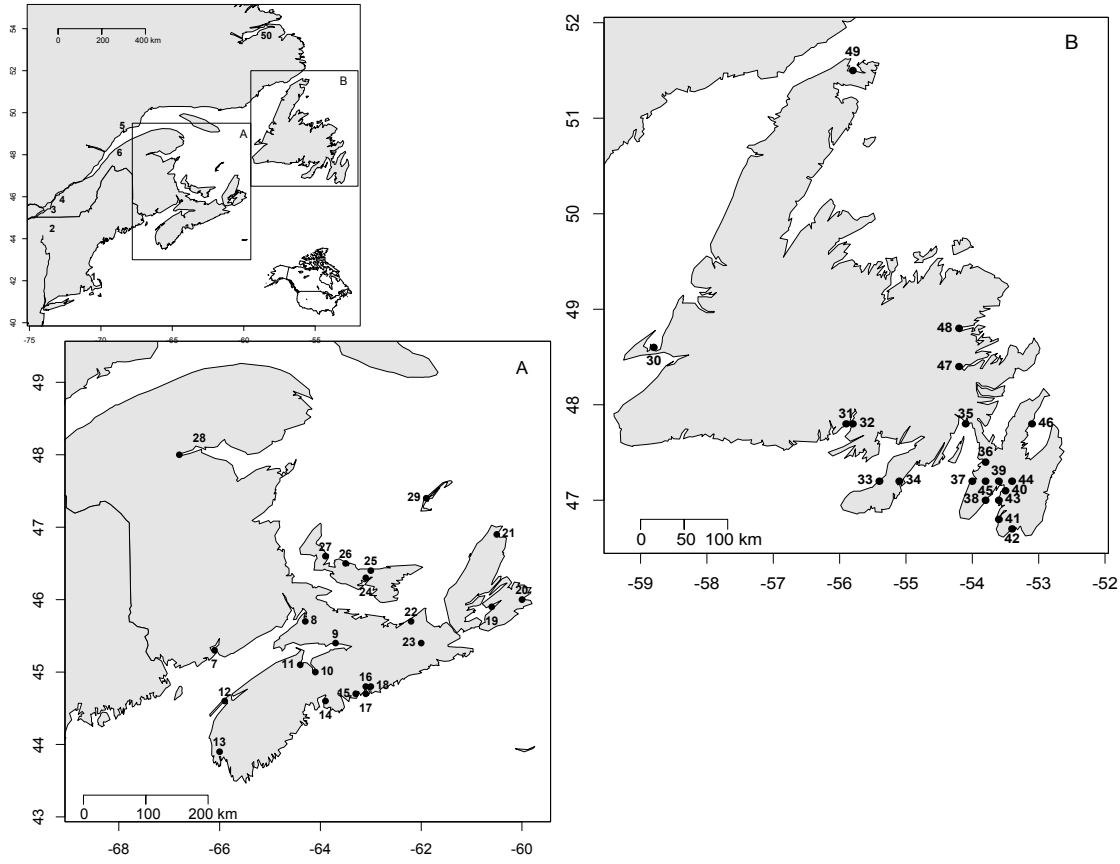

Figure 3.1 Map of the distribution of samples screened for either direct ND5 sequencing or RFLP analysis as part of the current study. Insets A and B are shown in closer detail for Newfoundland and the Maritime provinces, separately.

1991; Taylor & Bentzen 1993; Bernatchez & Martin 1996; Bernatchez 1997) was included in the overall analysis of haplogroups distribution. Samples comprised both anadromous and lacustrine life-history forms of rainbow smelt. Reference individuals from previous work for each of the identified races within the St. Lawrence estuary, were provided by J. Dodson (Université Laval, Québec, Canada).

3.2.2 DNA EXTRACTION AND NADH5 SEQUENCING

DNA was extracted using the protocol of Elphinstone *et al.* (2003), modified to work on a 96-well filter plate on a robotic liquid handling system (Perkin Elmer). PCR amplification of NADH subunit 5 (hereafter ND5) followed modified procedures outlined in Pigeon *et al*. (1998). Briefly, a 835 bp portion of the ND5 gene was amplified in 25 μL volumes as follows: 1 μL of template DNA, 2.5 μL of 10 × PCR buffer, 2.5 μL of 2 mM dNTPs, 0.75 μL of 10 μM forward and reverse primers (modified from Pigeon *et al*. 1998), and 1 U *Taq* DNA polymerase. PCR products were visualized on 0.8% agarose gels stained with Gel Green (Biotium, Inc.) and cleaned with the Exo-Sap protocol (USB Corporation) prior to sequencing. Briefly, 10 μL reactions contained 2 U of both ExoI and shrimp alkaline phosphatase (SAP) combined with 7 μL of PCR product in 1 × SAP buffer. Reactions were held at 37^{o}C for 30 min followed by enzyme denaturation at 80^{o}C for 15 min. Sequences were run on an ABI3730XL (Macrogen, Inc.) and were edited by eye using the program Sequencher (Gene Codes Corporation) and aligned using ClustalX v.1.81. An average of 10 individuals were sequenced from each of 30 locations for the initial survey (Table 3.1). Several samples were dropped due to significantly shorter read lengths from the DNA sequence analysis so as to maximize

the overall consensus length, which resulted in 723 bp of consensus sequences. However, these shorter sequences could still be categorized into either of the two races based on a minimum of two diagnostic sites and were therefore included in the overall frequency calculations of the two races.

After initial analysis and confirmation of the presence of the two glacial races based on direct DNA sequencing, a diagnostic RFLP test was developed to quickly assay further samples to estimate the relative frequency of the two clades in different geographical areas. To this end, a *Stu*I restriction site (recognition sequence AGGCCT) was utilized on the ND5 amplicon and resulted in either an uncut (835 bp; "B" race) fragment or a single cut site generating a 564 and 271 bp fragment ("A" race). The RFLP digest consisted of a 15 μL reaction containing 10 μL of ND5 PCR product, 1 × digestion buffer and 1 U *Stu*I enzyme (NEBiolabs). Digests were incubated at 37^{o}C for 60 min followed by enzyme deactivation at 80^{o}C for 20 min. Digested products were visualized on a 1% agarose gel stained with Gel Green (Biotium, Inc.) and scored as either "A" or "B". On average 18 individuals were surveyed from 41 sites by this method.

3.2.3 Statistical analyses

3.2.3.1 Descriptive Statistics of ND5 Sequence Data

DNA sequence data was analyzed according to two separate groupings: (1) location sampled and (2) major haplogroup (i.e. 'A' vs. 'B'; see Results). For each grouping, the number of polymorphic sites, haplotype and nucleotide diversities were calculated using DnaSP 5.10 (Librado & Rozas 2009). A median-joining haplotype network was constructed on the entire sequence dataset using the program NETWORK 4.6 (Fluxus Technology Ltd.). Each individual was then assigned to a haplogroup

identified by the network. An analysis of molecular variance (AMOVA) was then conducted using ARLEQUIN v.3.5.1.2 (Excoffier & Lischer 2010) on the individual sample locations, categorized as to the predominant haplogroup identified among individuals at each location and the variation was partitioned among haplogroups, among populations within haplogroups, and within populations.

3.2.3.2 Historical demographic patterns

The program ARLEQUIN was used to generate mismatch distributions (Rogers & Harpending 1992) on each of the major haplogroups, to test for evidence of a recent demographic expansion. The fit of the observed distribution was compared to the expected distribution under a model of sudden population expansion via 10,000 bootstrap replicates and assessed for goodness of fit via the sum of squared differences (SSD) and Harpending's raggedness index. An estimate of time of the expansion (in generations) can be obtained from the crest of the mismatch distribution τ via the equation $\tau = 2\mu t$, where μ is the mutation rate of the whole sequence under study and t is the time (in generations). Given a lack of fossil calibration for *Osmerus* or the substitution rate for the ND5 gene in smelt, a per-site mutation rate (calibrated to the 723 sites considered here) of 1% per million years was assumed based on an estimated mitochondrial molecular clock for salmoniformes (Smith 1992; cf. Bernatchez 1997), with an average generation time of two years for smelt, however this could be longer, especially in the northern part of their range (McKenzie 1964). In addition, Tajima's *D* (Tajima 1989) and Fu's *F*s (1997) tests of neutrality were calculated for each haplogroup, as these statistics are expected to yield significant deviations from expected stationary values under population expansion. Significance was assessed via 10,000 bootstrap replicates. I furthermore,

attempted to construct Bayesian Skyline Plots (BSP) (Drummond *et al.* 2005) in the program BEAST (Drummond & Rambaut 2007) as an alternative to the mismatch distributions for reconstructing historical population size changes through the coalescence. Unfortunately, the program failed to converge for runs of varying lengths (up to 5 x 10^8 iterations). This is potentially due to the low information content of the smelt lineages as each is dominated by a single haplotype with few rare variants. Therefore, results of this analysis will not be reported. A similar failure to converge was documented by Moore (2012) for Arctic charr with similar mismatch distributions as observed here (see Results).

3.2.3.3 ND5 RFLP Assay

Individuals screened via the PCR-RFLP assay were visualized and scored as either 'A' (cut) or 'B' (uncut) and are presented in Table 3.1. These individuals, along with the sequenced individuals from the current study and those taken from earlier studies were added, per location, to map the distribution of the two identified races.

3.3 RESULTS

3.3.1 MTDNA SEQUENCE VARIATION HAPLOTYPE DISTRIBUTION

Sequence data were generated for 1-19 individuals per location across a total of 30 locations. The number of haplotypes observed per location was 1-13. The highest haplotypic and nucleotide diversity was found at the Fore River (Boston) location, which was the most southerly location screened for sequence variation. Among these 13 haplotypes, individuals differed at 1-6 nucleotide positions, within the same major haplogroup. Eight sites had a single haplotype and with the exception of the Magdalen

Islands, all other locations were in Newfoundland, three of which were freshwater populations. Mainland samples had an overall higher level of haplotype and nucleotide diversity (0.550 and 0.0022, respectively) than samples from Newfoundland (0.220 and 0.0009, respectively). The Point Amal location (site #30, Figure 3.1; Table 3.1) had the highest level of diversity among all Newfoundland populations.

The haplotype network (Figure 3.2) revealed the presence of two major haplogroups, confirming earlier characterization of glacial races in this species. The two main haplotypes differed by four diagnostic sites and had an average sequence divergence of 0.6%, similar to previous estimates based on whole mtDNA RFLP analyses (Bernatchez 1997). The average sequence divergence within each lineage was 0.12% and 0.04% for the 'A' and 'B' types, respectively. Interestingly, three haplotypes were resolved that were intermediate in character between the two haplogroups. One of these, consisting of four individuals from the Lac Saint Pierre sample, differed by two mutations from either main haplotype in the two groups. The other two intermediate haplotypes were one and three steps away from the intermediate Lac Saint Pierre haplotype and both of these were found in individuals from the Fore River (Boston) sample. The presence of known reference samples from previously defined 'A' and 'B' races allowed the two haplogroups resolved here to be standardized to previous studies. Each individual was then characterized as either 'A' or 'B' type, and the overall location categorized as 'A' or 'B' depending upon the predominating type at that site.

An analysis of molecular variance, among samples for which DNA sequence data was available, revealed significant variation between groups (i.e. 'A' vs. 'B') and among populations within groups (Table 3.2). Each sample location was categorized as 'A' or 'B' depending on the predominating haplogroup at that site. While some sites contained

both lineages, most had one type present at ≥70% frequency (Figure 3.4), and any occurrence of both types at a site was included in the 'among populations within groups' hierarchy to take account of different haplotypes at a location. Genetic differences between the two races explained 69% of the variance ($\Phi_{CT} = 0.690$). In contrast, 1.76% of the variance was explained among populations within either group ($\Phi_{SC} = 0.057$) and 29% of the variation was attributed to variation within populations ($\Phi_{ST} = 0.708$).

Individuals subsequently screened for belonging to either the A or B clade via the RFLP assay were scored on the presence of a single fragment ('B') or two fragments ('A') and were included in the totals of the proportion of individuals for each haplogroup per location. In addition, information on the proportion of A vs. B haplotypes from previous studies (Baby *et al.* 1991; Taylor & Bentzen 1993; Bernatchez & Martin 1996, Bernatchez 1997) was incorporated to allow for a more comprehensive coverage of the natural distribution of rainbow smelt. For these latter locations, sites were chosen to complement the coverage from the current study. As previously reported, the occurrence of the B race was highest in the western portion of the distribution, throughout the Hudson River Valley and upper St. Lawrence, while the 'A' race predominated in the eastern portion of the distribution, covering much of the Maritime provinces of Canada. Additionally, the two races have been observed in a zone of secondary contact along the upper St. Lawrence River estuary. However, due to the increased geographical coverage presented here, a second zone of secondary contact between the lineages was observed in Newfoundland, with sites along the west and south coasts predominantly of type 'A'

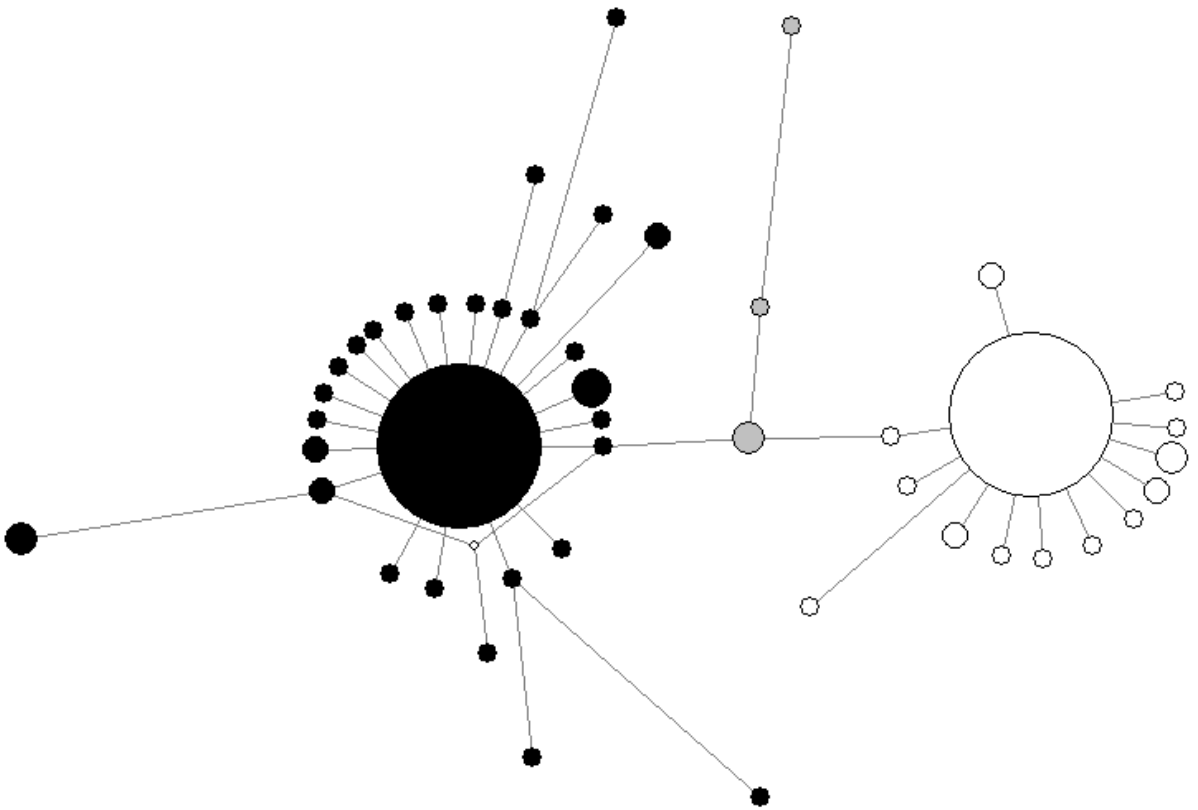

Figure 3.2 Haplotype network of 723 bp of ND5 mtDNA sequence from 261 rainbow smelt. Haplotypes in black represent the 'A' mtDNA type, while those in white represent the 'B' mtDNA type. Gray haplotypes are intermediate between the two main haplogroups.

Table 3.2 Results of the analysis of molecular variance (AMOVA) for genetic divergences in the ND5 mtDNA sequence data. Populations were partitioned into major haplogroups (A vs. B) depending upon the predominant group at each particular geographical location. Percentage of variance explained, fixation indices and significance of associated *P* values based on 10,000 permutations are included for each hierarchical level.

Source of variation	d.f.	Percentage of total variance	Fixation indices	*P* value
Among groups	1	69.02	Φ_{CT} = 0.690	<0.0001
Among populations within groups	24	1.76	Φ_{SC} = 0.057	<0.05
Within populations	231	29.22	Φ_{ST} = 0.708	<0.0001
Total	**256**			

while those on the east coast and Avalon Peninsula were predominantly of type 'B' (Figure 3.3). It is of interest that a single location on the Avalon Peninsula (Gull Pond) was exclusively type 'A' despite the surrounding locations predominantly or exclusively being composed of type 'B' (Figure 3.3). A similar disparity between neighbouring sites has been previously observed in Green Lake (Maine), which was exclusively of type 'B' but surrounded by locations of type 'A' (Figure 3.3; Taylor & Bentzen 1993). Among the Maritime Provinces, the occurrence of the 'B' type ranged between 0% (multiple sites) and 42% (Piper's Creek, PEI). Two locations sampled as part of the current survey (Mosher's and Steven's Brooks, Nova Scotia [sites 16 and 17, Figure 3.1]) demonstrate bimodal spawning runs (see Chapter 6) and samples from both runs were screened for the RFLP. It is of note that the two 'early' runs were all type 'A', while for the two 'late' runs between 9%-13% of the samples were of type 'B'. Given the small sample sizes, however, it may be of interest to screen further individuals to expand upon this observation.

Following from the spatial patterns described above, the frequencies of the A and B haplotypes vary longitudinally with the 'B' race predominating in the western and eastern most portions of the range surveyed and the 'A' type most predominant throughout the central region (Figure 3.4). As the samples considered comprise both anadromous and lacustrine populations of smelt, Figure 3.4 depicts the two life-history forms and plots the proportion of the 'B' type against longitude. In total, 24 lacustrine samples were considered (most from earlier studies), of which 19 (79%) were fixed for either type and not always of the same type as surrounding locations (Figure 3.2, Figure 3.4).

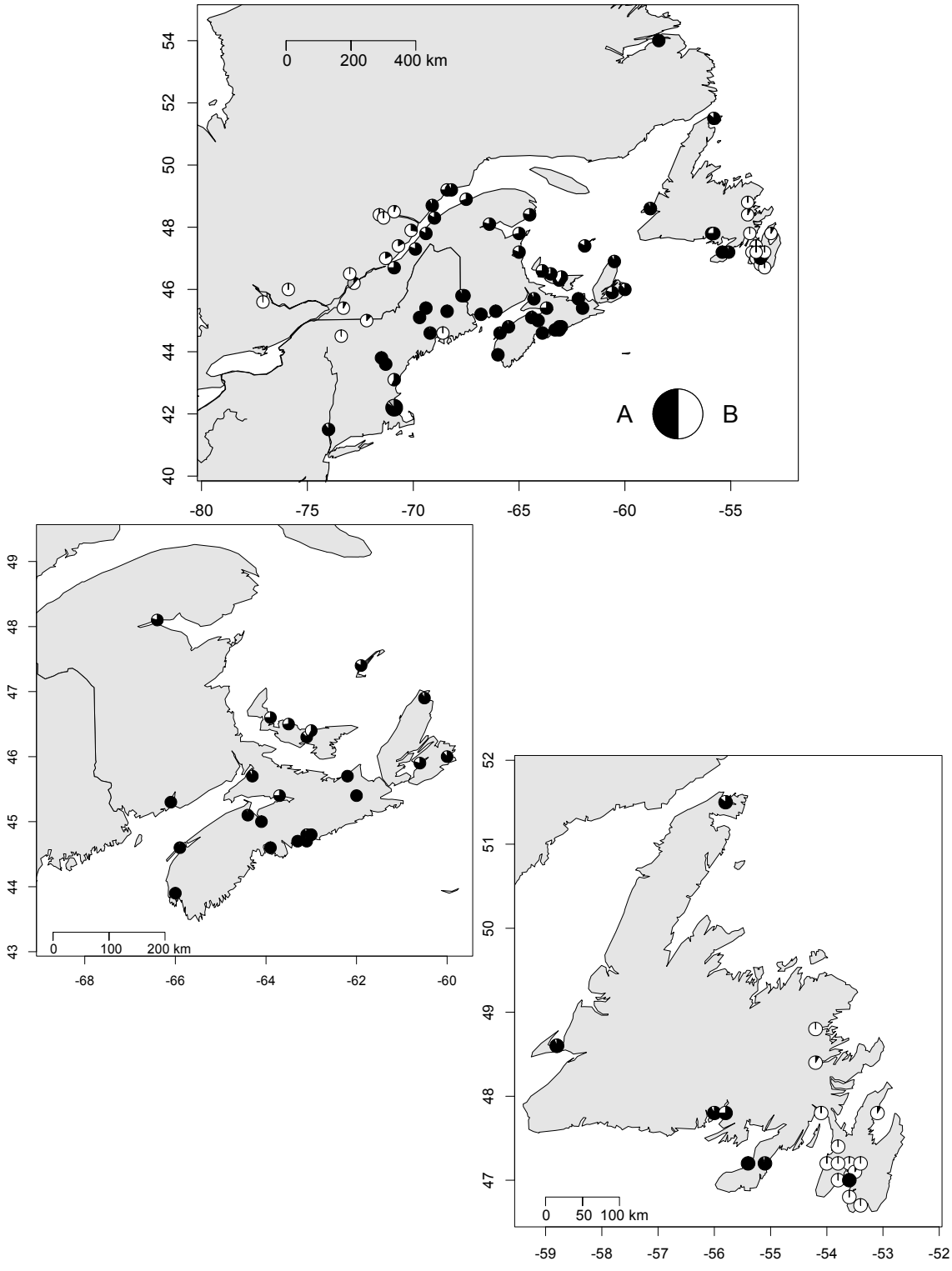

Figure 3.3 Distribution of haplogroups ‘A’ and ‘B’. The overall map on the top includes samples summarized in Bernatchez (1997), while those below depict only sites screened for the current study.

3.3.2 HISTORICAL DEMOGRAPHIC PATTERNS

The mismatch distributions for both A and B races were unimodal (Figure 3.5), and had a non-significant SSD and raggedness index, consistent with a pattern of demographic expansion. The tau value (τ; Table 3.3), which reflects the location of the mismatch crest and its distribution, provides a rough estimate of the time when expansion started. The observed values of the age of expansion (τ) were 0.7 and 3.0 for races A and B, respectively, suggesting a more recent population expansion for the former. The estimate of the population expansion for the two races, based upon a 1% mutation rate per site per million years was about 24,000 YBP and 103,000 YBP for A and B types, respectively. In addition, both Tajima's *D* and Fu's F_S neutrality tests were statistically significant (Table 3.3) for both lineages, supporting a conclusion of recent demographic expansion.

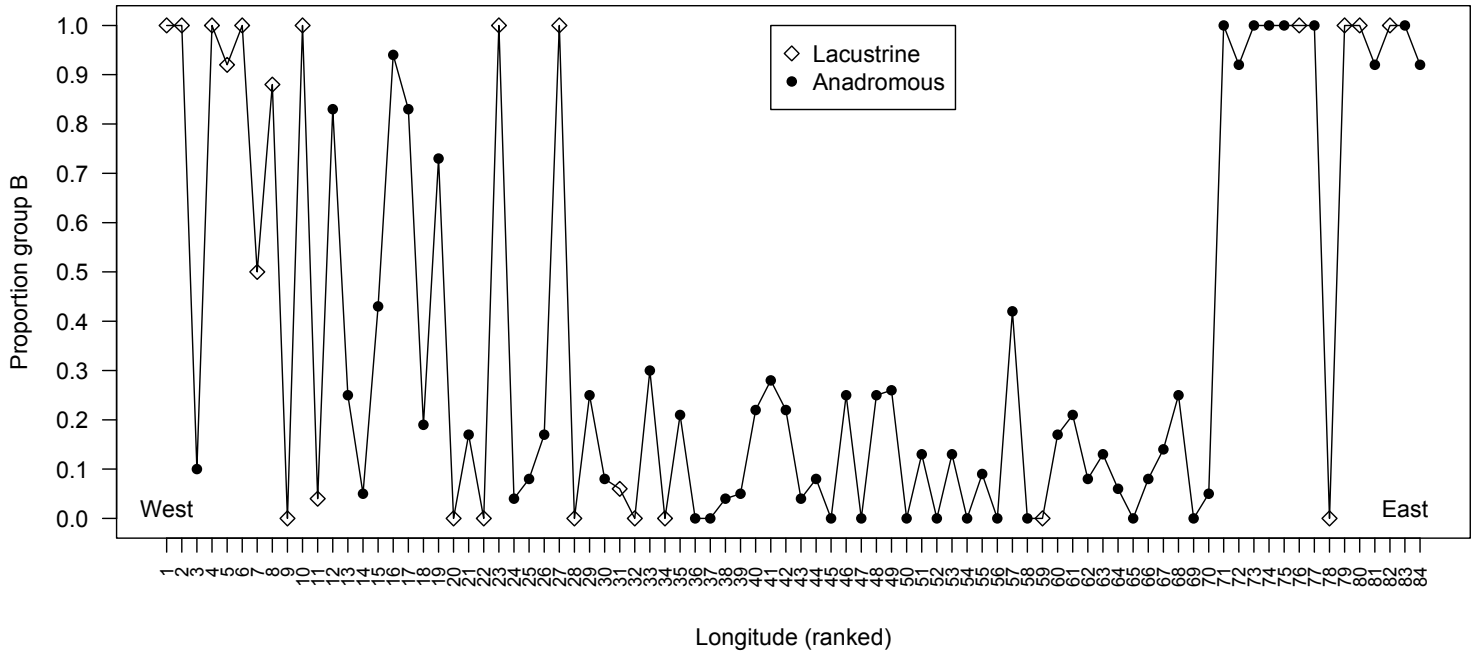

Figure 3.4 Frequency distribution of the 'B' mtDNA type among all smelt populations ranked by longitude, east to west and identified by either lacustrine or anadromous life-history type.

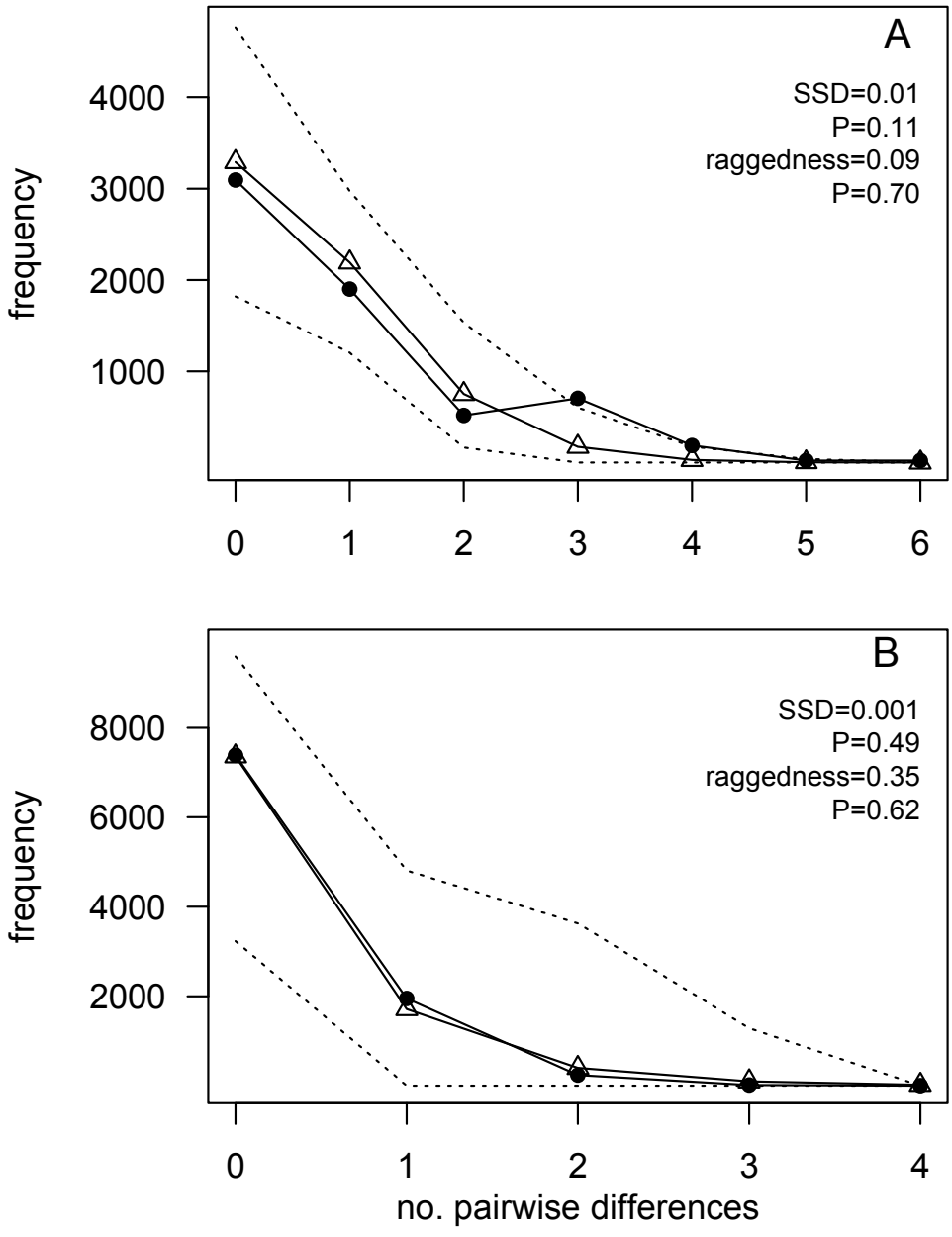

Figure 3.5 Mismatch distributions for 'A' and 'B' mtDNA types. Circles represent the observed mismatch distribution while open triangles represent the expected distribution under a model of demographic expansion. Dotted lines represent the 95% CI based upon 10,000 replicates.

Table 3.3 Genetic diversity measures for each of the two identified mtDNA types along with parameter estimates of population expansion (tau).

	A	**B**
Number of sequences	114	139
Number of haplotypes	28	13
Haplotype diversity (*h*)	0.52 ± 0.06	0.23 ± 0.05
Nucleotide diversity (π)	0.0013 ± 0.0024	0.0004 ± 0.0012
Tajima's *D*	-2.55	-2.30
P-value	<0.0001	<0.0001
Fu's F_S	-32.23	-19.80
P-value	<0.0001	<0.0001
tau (τ)	0.701	3.00
Lower tau	0.000	0.389
Upper tau	1.45	3.5

3.4 DISCUSSION

3.4.1 PATTERNS OF DISTRIBUTION OF TWO GLACIAL RACES OF SMELT

Direct mtDNA sequencing and increased geographical coverage corroborated the existence of two distinct phylogenetic groups of smelt throughout its native range. These two groups have been acknowledged for some time, and this dichotomy has been shown to be independent of both life-history type (i.e. lacustrine or anadromous) and trophic ecotypes within landlocked populations (i.e. macro- vs. microphagous) (Taylor & Bentzen 1993a; Bernatchez 1997), although morphological differences have been associated with both trophic specialization and proposed vicariant events (Taylor & Bentzen 1993b; Barrette *et al.* 2009). The main finding here, however, was the existence of a second zone of secondary contact. While the two races have been documented to occur in relatively close proximity on opposite shores of the St. Lawrence estuary, this work demonstrates a transition zone between the races in eastern vs. southern and western Newfoundland. This finding presents an unusual pattern whereby one of two races predominates towards opposite longitudinal ends of a species' range, with the other form most common at intermediate longitudes.

Despite relatively small sample sizes per location for the direct mtDNA sequence data (see Table 3.1), populations in Newfoundland generally displayed lower levels of haplotypic and nucleotide variation compared to the mainland. Lower levels of genetic variation of smelt in Newfoundland have also been demonstrated with nuclear markers (see Chapter 5; Bradbury *et al.* 2006) compared to the mainland portion of the species range. The exception is the southwestern Newfoundland population at Point Amal, which had a high level of mtDNA haplotype variation comparable to that seen in mainland

populations, again a pattern also observed with nuclear variation. The occurrence of this pattern with both types of markers suggests that this region of Newfoundland has experienced greater levels of historical gene flow with mainland populations, and may represent the first area of colonization of the island, at least for smelt bearing type 'A' mtDNA. Such a colonization route would seem plausible as the southwest corner of Newfoundland is the closest geographically to mainland source populations.

As a result of these findings of two glacial races in Newfoundland, Bradbury *et al.* (2011) conducted a combined study of mtDNA and microsatellites and found similar trends between marker types in the distinction of southeast Newfoundland from other Newfoundland populations. Notably, a single microsatellite, *Omo*11, had two predominant alleles, one largely associated with the Avalon Peninsula, which coincided with the presence of the 'B' mtDNA type. However, even excluding this particular microsatellite locus, the distinction remained. While similar trends between marker types were apparent, they were not identical. Microsatellite alleles typically associated with the 'B' race were found in the west and north coasts and most evident was the discrepancy surrounding the transition zone from the Avalon Peninsula going north and west. The nuclear markers suggested a secondary contact zone in Conception Bay to Bonavista Bay to the north, whereas the mtDNA presented here suggested a contact zone much further north in an area not sampled and demonstrated a more east-west split in Newfoundland (see Bradbury *et al.* 2011). Furthermore, the finding of the 'A' mtDNA type in Gull Pond surrounded by 'B' mtDNA types is striking and mirrored in a similar observation for Green Lake (but for the opposite mtDNA type; see Taylor & Bentzen 1993; Bernatchez 1997). Such cyto-nuclear discordance is not uncommon and many potential explanations can be invoked. Factors such as sex-biased dispersal, asymmetric introgression, a four-

fold smaller N_e (and hence greater drift) for mtDNA compared to nuclear markers, and differential levels of selection between marker types have all been suggested (see Toews & Brelsford 2010 for a review). While sex biased dispersal is common in nature and has indeed been documented in salmonids (Hutchings & Gerber 2002), little about this is known for smelt. As with many marker types, it may not be safe to assume neutrality, as has historically been the case for mtDNA (Ballard & Whitlock 2004). Bradbury *et al.* (2011) estimated the strength of selection required to maintain the observed clines in Newfoundland for both the mtDNA and Omo11 as <0.01, consistent with other values in the literature. They showed that relatively weak selection could act to reinforce the observed gradients in haplotype or allele frequencies.

Given the aforementioned patterns of genetic variation and diversity in Newfoundland and associated patterns of cyto-nuclear discordance (Bradbury *et al.* 2011), a scenario of differential introgression may be considered. For instance, most Newfoundland smelt populations of type 'A' may be the result of introgression between type 'B' (presuming a Grand Banks origin; see below) and type 'A' smelt colonized from the mainland. Type 'A' mitochondrial variants may have drifted to near fixation, but some nuclear genes may derive from the type 'B' race. This could arise due to founder effects and differential selection for mtDNA types as mentioned above or alternatively different random outcomes for unlinked genetic markers. Fixation of mtDNA types against a mixed nuclear background could follow, for instance from asymmetric mate choice between races, which has been documented from such discordance in hybrid zones (Asmussen *et al.* 1989; April & Turgeon 2006).

Interestingly, the consistent pattern of occurrence of the type 'B' mtDNA and nuclear alleles on the Avalon Peninsula may be due, in part, to the later colonization of

this area as it was glaciated for a longer period of time than the rest of Newfoundland (see Shaw *et al.* 2006). Furthermore, it has long been recognized that the ice cap on the Avalon Peninsula was independent of the one that covered the rest of Newfoundland during the late Wisconsinan, and deglaciation of this ice cap began *c*. 10,000 YBP (Coleman 1926; Henderson 1959; Cato 1998). Therefore, less time and opportunity may have been available to allow for the dissociation of cyto-nuclear variation in this region.

For the mainland sites surveyed here, results were consistent with earlier conclusions that in the east the 'A' type predominated with generally increasing levels of the 'B' type westward toward the St. Lawrence and Hudson River valleys (Bernatchez 1997). However, several other observations made here are of interest and suggest the possibility for further work. Firstly, there is the occurrence of haplotypes in Lac St. Pierre and the Fore River, that were genetically intermediate between the two races. This suggests that the southerly portion of the range is where the initial divergence occurred and still acts as a reservoir of diversity maintained through the various glacial cycles. Given that only two sites sampled here contained these intermediate haplotypes, and the relatively small sample sizes for the sequence data, it would be of interest to extend the sample sizes and add further locations in the south for sequence analysis to explore the extent of the signature of initial divergence and the depth of variation retained in these areas.

The second point of interest was the difference in haplotype frequencies between 'early' and 'late' spawning runs in Musquodoboit Harbour (Mosher's and Steven's Brooks) on the Atlantic coast of Nova Scotia. The occurrence of bimodal spawning runs in each of these brooks has been suggested to have arisen *in situ*, influenced at least partly by high egg density and mortality resulting in a disruption of spawning activity

(Coulson *et al.* 2006; see Chapter 6). However, the occurrence of the 'B' mtDNA type at a frequency of 9-13% in the 'late' runs, while absent in the 'early' runs suggests the possibility of a secondary contact between runs of differing mtDNA types. Furthermore, three other sites were sampled along the Atlantic coast, and in location all individuals were of the 'A' mtDNA type. This suggests that the 'B' type is only associated with the 'late' runs in this region and further sampling is therefore warranted.

3.4.2 Demography and Postglacial Recolonization

Both 'A' and 'B' mtDNA types showed evidence supporting recent population expansions, as have been reported in a number of species associated with past vicariant events (e.g. Atarhouch *et al.* 2006; Zane *et al.* 2006; Petersen 2007; Liu *et al.* 2007; Hubert *et al.* 2007; Liu *et al.* 2011). However, in both cases the timing of the expansion pre-dates the last glacial maximum with the expansion of the 'B' mtDNA type (~103,000 YBP) estimated as approximately five times older than the more recent expansion of the 'A' mtDNA type (~24,000 YBP). However, applying the "standard clock" of 2% divergence per million years for mtDNA (Brown *et al.* 1982), would yield an estimate of expansion for the 'A' mtDNA type of smelt ~12,000 YBP, thereby during the deglaciation of the species range, while that for the 'B' mtDNA type would still predate the last glacial maximum (~50,000 YBP). Dodson *et al.* 2007 inferred expansion of Arctic and northwest Atlantic capelin (*Mallotus villosus*, a close relative of *Osmerus*) coinciding with the warm Sangamon interglacial that occurred 135,000-115,000 YBP, while other capelin lineages were inferred to have under gone expansion during even earlier interglacial periods (200,000-300,000 YBP). They concluded that range and demographic contractions caused by the most recent glaciations did not appear to have

influenced the genetic architecture of their capelin populations. Bernatchez (1997) reached a similar conclusion regarding rainbow smelt due to the higher diversity found within each of the mtDNA types compared to other north temperate fishes. Given the uncertainty surrounding estimates of substitution rates and their subsequent impact on demographic or evolutionary estimates, coupled with situations where a lineage specific or gene specific calibration is not available, the estimates presented here do not permit a robust dating of expansion. However, it is possible and useful to make deductions about the *relative* patterns and timings inferred from nucleotide change within a species (Gharrett *et al.* 2001). Therefore, regardless of the exact dates of population expansion for the two mtDNA types of smelt, the 'A' lineage appears to have a much more recent expansion.

Bernatchez (1997) estimated the divergence of the two clades at ~700,000 YBP. Based upon the sequence divergence for this portion of the ND5 gene and assuming the same 1% divergence per million years, the estimate derived here of ~550,000 years is consistent. Such a timeframe implies that the divergence of the two smelt lineages certainly predates the Wisconsinan glaciation. Given the intermediate haplotypes found in locations associated with the Hudson River and St. Lawrence valleys, it seems reasonable to assume that the initial divergence occurred in the southerly portion of the species distribution, likely associated with earlier vicariant events.

Prior to the discovery of the 'B' mtDNA type in Newfoundland, the prevailing recolonization scenario was that 'A' mtDNA type had colonized much of the east, presumably from a Grand Banks refugium off the coast of Newfoundland. The 'B' mtDNA type was attributed to an Atlantic coastal plains refugium south of the ice sheet. Colonization from these two sources was inferred to have led to the observed secondary

contact in the upper St. Lawrence estuary (Bernatchez 1997). The occurrence of the 'B' mtDNA type in eastern Newfoundland suggests that this scenario needs to be revised.

The close proximity of the 'B' type populations to the Grand Banks and near complete fixation of this variant, suggests that this lineage survived the last glaciation on the exundated Grand Banks. As mentioned earlier, the Avalon Peninsula is thought to have remained ice-covered longer than the rest of Newfoundland (Shaw *et al.* 2006) and the close coupling of the 'B' mtDNA type and associated nuclear alleles present (Bradbury *et al.* 2011) would suggest a more recent colonization. Several instances of secondary contact between eastern and western races, such as that for Atlantic salmon have been reported off Newfoundland (Verspoor 2005).

Given the Grand Banks hypothesis for the 'B' mtDNA type (rather than 'A' as previously described; Bernatchez 1997), the likely last glacial refugium for the 'A' mtDNA type is thought to be the Atlantic coastal plains, south of Cape Cod, Massachusetts, which was also associated with George's Bank off southwest Nova Scotia, Canada (Shaw & Courtney, 2002) and has been suggested as a refugium for a number of fish species (Curry 2007). The possibility of Sable Island providing a refuge has been debated but more recent studies suggest this area was completely covered by glaciers (Dyke *et al.* 2002). Given the lack of natural recolonization of smelt into the Great Lakes, Bernatchez (1997) ruled out the possibility of a Mississippian refuge.

Regardless of the exact location of refugia for rainbow smelt, the predominance of the 'B' mtDNA type at both the western and eastern ends of the species' range, with the 'A' type predominating at mid-longitudes is intriguing. Such a pattern may at first appear difficult to interpret with regards to how recolonization would have occurred. However, given the repeated series of range and demographic contractions and

expansions experienced by many species over long periods of time, the genetic signal and distribution of lineages from previous vicariant events are likely to be eroded and/or changed over time. The observation of a much older population expansion for the 'B' mtDNA type may suggest that this lineage was previously more widespread and predominated throughout a larger geographical extent than at present. It is possible that upon glacial advance the 'B' race was sequestered in both the Atlantic coastal plains and the Grand Banks. Upon deglaciation, this lineage would have recolonized the Avalon Peninsula and the North Shore of the St. Lawrence estuary via the post-glacial Champlain Sea. The 'A' race could have persisted on the exundated George's Bank and recolonized much of the Maritime provinces and eastern Quebec and Maine. The more recent expansion evidenced by the mismatch distribution for the 'A' race suggests that this race may have been geographically (and demographically) more limited previously. It is also noteworthy the 'B' race occurs along the north shore of the St. Lawrence and the Avalon Peninsula, both regions that have been implicated in remaining ice-covered a longer period of time (e.g., Dubois & Dionne 1985, Shaw *et al.* 2006).

Regardless of the exact locations of glacial refugia for the different lineages of smelt, the current study demonstrates the importance of comprehensive geographical coverage, spanning the entire distribution of the species. Inadequate sampling in Newfoundland in earlier studies led to conclusions about the glacial history and post-glacial colonization of rainbow smelt that have been shown to be implausible in the light of the more comprehensive data collected in this study. These data support a more complex interpretation of the glacial and post-glacial history of rainbow smelt and sheds light into the patterns associated with historical events in shaping the genetic diversity and differentiation within a species.

CHAPTER 4

CHARACTERIZATION OF DI- AND TETRANUCLEOTIDE MICROSATELLITE MARKERS IN RAINBOW SMELT (*OSMERUS MORDAX*)

4.1 INTRODUCTION

The rainbow smelt (*Osmerus mordax*) is a small North American osmerid fish ranging from New Jersey north to Labrador (Scott & Scott 1988). Across this range the species is an important forage species for predatory fishes, birds and mammals, and is also the target of recreational and commercial fisheries. Smelt are typically anadromous but landlocked, completely freshwater populations have been described (Kendall 1927; McAllister 1963; Nellbring 1989). Within the freshwater environment two ecotypes can occur as both allopatric and sympatric populations: a 'dwarf' form and a 'normal' form that more closely resembles the anadromous ecotype. The freshwater dwarf and normal smelt ecotypes differ in a number of features including size at age and gill raker counts (Copeman 1977; Lanteigne & McAllister 1983; Taylor & Bentzen 1993a). Differences between the ecotypes have also been investigated with genetic analyses (Taylor & Bentzen 1993a,b; Saint-Laurent *et al.* 2003). Given the diversity of environments in which they occur and the differing life history strategies exhibited, smelt offer potential as a model species for the study of a number of evolutionary and population genetic questions. To date, most of the genetic work has focused on mitochondrial DNA (mtDNA) (Taylor & Bentzen 1993a; Baby *et al.* 1991; Bernatchez 1997). Here, we describe and characterize 12 novel microsatellite loci, including 10 tetranucleotide loci and two compound dinucleotide-tetranucleotide microsatellites, developed as part of a population genetic assessment of rainbow smelt.

4.2 METHODS/RESULTS

Genomic DNA from one anadromous rainbow smelt was used to create microsatellite-enriched libraries for (GACA). DNA was restriction digested with *Rsa*I or *Hin*cII and ligated to SNX linkers (see Hamilton *et al.* 1999). The linker-ligated fragments were polymerase chain reaction (PCR) amplified using the SNX linkers as the priming sites and enriched for microsatellites by hybridization to biotin labelled $(GACA)_4$ probes and collected using streptavidin-coated magnetic beads (Dynal). The microsatellite-enriched library was PCR amplified and cloned into Invitrogen MaxEfficiency competent cells. Positive colonies were screened for suitably sized inserts (300-1000 bp) by direct PCR amplification of colony picks using M13 sequencing primers under standard PCR conditions and imaged with agarose electrophoresis. Plasmid DNAs were isolated using a QIAGEN plasmid kit and sequenced on a CEQ 8000 (Beckman Coulter). PRIMER3 software (Rozen & Skaletsky 2000) was used for primer design.

We tested the 12 loci using DNA from a typical anadromous population, the Portapique River, Nova Scotia. Fish were collected by dip netting. Pectoral or caudal fin clips were taken and immediately placed in 95% ethanol. DNA was extracted following the protocol of Elphinstone *et al.* (2003), modified to work with a 96-well filter plate and automated on a robotic liquid handling system (Perkin Elmer). Individuals were genotyped using PCR conditions of 5- or 10-µL volumes containing 20-100 ng DNA, 1.5 mM $MgCl_2$, 80 µM each dNTP, 0.5 U *Taq* DNA polymerase (Applied Biosystems), 0.3 µM each primer (forward primers were end-labelled with Hex dye), and 1x PCR buffer. We used touchdown PCR for all but one of the primer pairs (Omo6) which used the

conditions as follows: 95°C for 3 min, followed by 30 cycles of 95°C for 30 s, annealing temperature of 65°C for 30 s and 72°C for 30 s. This was followed by a final extension at 72°C for 3 minutes. Two temperature profiles were used for touchdown PCRs of the remaining 11 primer pairs (OmoTD2 and OmoTD5), in order to allow for the possibility of multiplex PCRs. Touchdown PCR conditions were as follows: 94°C for 2 min, followed by 4-5 cycles of 94°C for 30 s, program specific touchdown annealing temperatures (T_a) (Table 1) minus 1°C per cycle for 30 s, 72°C for 30 s, followed by 25-26 cycles where the T_a was held constant at 4°C below the starting temperature. The remainder of the steps were as reported for the touchdown phase of the program. A final extension was held at 72°C for 5 min. Reactions were run on Eppendorf thermocyclers and imaged on an FMBioII system (Hitachi Genetic Systems).

We calculated observed and expected heterozygosities using Genetix (version 4.05.2, Belkhir *et al.* 2004) and tested for Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium using FSTAT (version 2.9.3.3, Goudet 2002). No departures from HWE were detected (data not shown). There was evidence of linkage between Omo3 and Omo16 ($P = 0.00076$). The microsatellite loci presented here are being used to examine population structuring among natural smelt populations.

Table 4.1 Summary characteristics for 12 smelt primers with locus identifiers, the repeat motif and primer sequences including annealing temperature (T_a), number of alleles, the number of individuals genotyped at that allele (n), the allelic size range and observed (H_O) and expected heterozygosity (H_E).

Locus GenBank no.	Repeat motif	Primer sequence (5' – 3')	T_a(°C)*	No. alleles (n)	Size range (bp)	H_O	H_E
MGPL-Omo1	$(GACA)_{12}$	F: cggtcacgcaactaacatct R: cggctggttggctgtttat	66-62TD	15 (93)	104-162	0.7634	0.7327
MGPL-Omo2	$(GTCT)_8$	F: tgatgatggatgaggagctg R: caagcagtcagtcaggcaga	66-62TD	11 (83)	156-216	0.8072	0.8365
MGPL-Omo3	$(CTGC)_{19}$ctgttt$(GTCT)_{10}$	F: ggatttgccatgttgaagcta R: cacatgcacaacacagtcca	66-62TD	18 (84)	174-246	0.9048	0.9228
MGPLOmo4	$(CAGA)_{11}$	F: cagcgtctcaaatcacctca R: ctgcctgtctcccttactgc	66-62TD	10 (91)	170-222	0.7802	0.8055
MGPL-Omo5	$(CTGT)_9$ccgt$(CTGT)_2$	F: ctatgtgaacagaagctgtgaagag R: taaagacacctgccgacttg	66-62TD	15 (79)	226-282	0.8987	0.8753
MGPL-Omo6	$(CAGA)_{30}$	F: actgttggagatgaggcaca R: tctgtctgccttcctgtctg	65	9 (90)	168-206	0.8111	0.7309
MGPL-Omo9	$(GACA)_8$cagaca$(GGCA)_6$	F: tagcaggatcagcagggaga R: ttgcctgcctgtctctcttt	64-60TD	20 (86)	134-218	0.9186	0.9196
MGPL-Omo11	$(TGTC)_4(TGCCTGTC)_7$	F: ccttgaggcactgaaccact R: acatgcacatgcaggtaagg	64-60TD	16 (93)	134-214	0.8495	0.8586
MGPL-Omo13	$(TG)_{11}(TCTG)_9(TG)_4$	F: gccgctcgtaataatgctgt R: agggtagggaaagcgaaaga	64-60TD	22 (88)	152-202	0.9091	0.9277
MGPL-Omo14	$(CAGA)_{10}$catccagc$(CAGA)_{11}$	F: tgctgtgagaccttgttgatg R: aggtctgggatgaggagctg	64-60TD	24 (93)	186-322	0.9247	0.9440
MGPL-Omo15	$(GTCT)_7(GT)_{10}$ct$(GT)_8$	F: cccttcagcaatccttcaaa R: gggtaaatagtggagctgggta	64-60TD	19 (90)	140-242	0.8556	0.8422
MGPL-Omo16	$(CCTG)_{18}N_{20}(TCTG)_7$	F: tgaagctacaatcggtccatc R: cacatgcacaacacagtcca	64-60TD	19 (90)	162-234	0.8778	0.9241

*TD indicates a touchdown PCR.

CHAPTER 5

CONTRASTING SPATIAL SCALES OF GENETIC STRUCTURE AMONG ANADROMOUS RAINBOW SMELT

5.1 INTRODUCTION

Determining the extent to which populations of a species are connected via gene flow not only sheds light on the mechanisms and processes influencing population dynamics and species evolution, but can also play a major role in their effective management and understanding of their potential to adapt to future change. The preservation of evolutionary potential for a species through the designation of conservation units, requires the identification of population boundaries (Allendorf 1986). These boundaries are regulated, in part, by the level of gene flow among demes and the relative fitness of immigrants into populations and are key factors regulating population persistence. In some cases, obvious physical, geographic, or landscape features may restrict dispersal. Typically, however, for many marine species, the apparent lack physical barriers and prolonged larval stages have been associated with apparent high degrees of connectivity among demes. Indeed, isolation by distance (IBD) (e.g., Rousset 1997; Palumbi 2003) predicts that genetic differences accumulate along increasing geographic distances separating populations. However, local factors may also play a role in structuring populations on both small and large scales. On the other hand, differences associated with local adaptation in behaviour, morphology, physiology and other factors result in decreased fitness among non-native versus native individuals. The scale of population connectivity is therefore influenced both by the heterogeneity of the habitat as well as the biological characteristics of the species. For many aquatic species, factors

such as variation in the hydrographic and temperature conditions encountered throughout their range may therefore be a more pronounced determinant of population structure than increasing distance alone. Indeed, member-vagrancy has been proposed for many marine or anadromous species whereby the spatial scale across which early life history stages are retained determines the scale at which structuring occurs (Iles and Sinclair 1982).

The use of broad-scale sampling to infer the scale of population structure is crucial as patterns of differentiation even within a species may differ markedly (e.g. Bradbury *et al.* 2006), with historical vicariant events increasingly being implicated (e.g., Duvernell *et al.* 2008; Bradbury *et al.* 2011; Hasselman *et al.* 2013). Not only does the identification of such regional differences in population structure contribute to more efficient management of the species in question, but it aids in addressing questions regarding the factors that promote intraspecific differentiation, ultimately leading to speciation.

Rainbow smelt, *Osmerus mordax,* are assumed to remain near the coast and do not undergo lengthy migrations, suggesting a potential for strong population structuring. It has been shown that regional forms may exhibit strikingly different degrees of genetic structure. While Newfoundland populations display a strong signal of IBD, mainland populations appear to be much more highly connected (Bradbury *et al.* 2006, Kovach *et al.* 2013). Therefore it is possible that factors other than distance alone primarily determine the extent and scale of genetic structuring among mainland smelt. However, the occurrence of two distinct glacial race mitochondrial lineages in Newfoundland (Chapter 3; Bradbury *et al.* 2011) compared to a single race predominating throughout the Canadian Maritimes adds further complexity to regional comparisons.

This chapter expands upon earlier work (Bradbury *et al.* 2006) to explore patterns of genetic diversity and structuring across much of the native range of anadromous rainbow smelt, focusing on possible mechanisms underlying regional variation and in light of the recent findings of two mtDNA types in Newfoundland (Chapter 3; Bradbury *et al.* 2011).

5.2 Materials and Methods

5.2.1 Study Area

Anadromous runs of rainbow smelt were sampled between April and June during 2003-2006 from 61 populations by dipnetting, gillnetting or angling throughout much of their native range (Table 5.1; Figure 5.1) with, on average 75 individuals sampled per spawning run (range = 40-96). Pectoral or caudal fin clips were placed in 95% ethanol until DNA extraction. The range encompassed here spans between 42^{o}N – 54^{o}N latitude and contains both large river systems (e.g. Miramichi River, NB) to small brooks (e.g. Mosher's Brook, Musquodoboit Harbour, NS). The Cabot Strait separates Newfoundland from the mainland portion of the species range and is approximately 110 km across and 500 m deep. Within Newfoundland, the Labrador Current dominates coastal areas and the coastline is characterized by numerous large embayments. The mainland portion of the range contains a varied landscape with numerous bays and streams along the northeastern USA and the Maritime provinces and includes the Gulf of St. Lawrence and St. Lawrence River as well as the large tidal cycles in the Bay of Fundy.

Table 5.1 Locations sampled in the current study with sample sizes (N) and associated genetic diversity statistics (H_e, expected heterozygosity; H_o, observed heterozygosity, F_{IS}, inbreeding coefficient; A_r, allelic richness). Code refers to sample points in Figure 5.1.

Code	Site	Location	N	H_e	H_o	F_{IS}	Mean no. alleles	A_r
1	Fore	Massachussets Bay	93	0.854	0.860	-0.002	14.38	12.130
2	Parker	Ipswich Bay	93	0.864	0.875	-0.007	15.00	12.500
3	Kennebecasis	St. John River	74	0.818	0.772	0.065	12.25	10.940
4	Maccan	Cobequid Bay	40	0.832	0.859	-0.020	13.00	12.620
5	Portapique	Minas Basin	92	0.844	0.842	0.009	15.13	12.360
6	Kennetcook	Minas Basin	91	0.839	0.838	0.007	14.00	11.750
7	Gaspereau	Minas Basin	91	0.847	0.848	0.005	14.38	12.060
8	Haight	St. Mary's Bay	48	0.839	0.870	-0.025	11.75	11.080
9	Bingays	St. Mary's Bay	43	0.828	0.865	-0.032	11.88	11.450
10	Spechts	St. Mary's Bay	50	0.837	0.847	-0.002	12.38	11.390
11	Youngs	St. Mary's Bay	91	0.834	0.826	0.016	13.13	11.240
12	Tusket	Yarmouth	93	0.814	0.824	-0.006	12.25	10.140
13	FoxPoint	St. Margaret's Bay	44	0.759	0.785	-0.023	9.63	9.280
14	Boutilier	St. Margaret's Bay	93	0.788	0.803	-0.013	10.75	9.710
15	GlenHaven	St. Margaret's Bay	50	0.813	0.846	-0.030	10.50	10.110
16	Smelt Cove	Porter's Lake	93	0.806	0.804	0.008	12.25	10.470
17	MosherEarly	Musquodoboit Harbour	83	0.799	0.765	0.050	12.13	10.560
18	StevensEarly	Musquodoboit Harbour	91	0.793	0.817	-0.024	11.63	10.110
19	Smelt Brook	Jeddore Harbour	82	0.790	0.778	0.022	11.63	10.100
20	Malagawatch	Bras d'Or Lakes	47	0.817	0.831	-0.005	11.63	11.040
21	Dennys	Bras d'Or Lakes	54	0.835	0.816	0.034	13.13	12.230
22	Washabuck	Bras d'Or Lakes	43	0.824	0.808	0.032	11.25	10.750
23	Eskasoni	Bras d'Or Lakes	52	0.833	0.840	0.002	12.25	11.380
24	Mira	Cape Breton	50	0.823	0.861	-0.036	11.50	10.870
25	NHrbCB	Cape Breton	42	0.819	0.819	0.014	10.88	10.720
26	Doctors	Gulf of St. Lawrence	54	0.825	0.827	0.007	11.63	10.820
27	Knoydart	Gulf of St. Lawrence	63	0.836	0.862	-0.022	13.13	11.330
28	Newton	Hillsborough Bay	58	0.846	0.819	0.040	13.75	12.360

Table 5.1 *continued*

Code	Site	Location	N	H_e	H_o	F_{IS}	Mean no. alleles	A_r
29	Wrights	Hillsborough Bay	89	0.839	0.830	0.016	13.75	11.380
30	NorthRiver	Hillsborough Bay	90	0.851	0.834	0.026	13.13	11.570
31	Tannery	Hillsborough Bay	45	0.829	0.817	0.027	12.38	11.940
32	DouglasBr	Hillsborough Bay	86	0.845	0.824	0.031	13.75	11.750
33	Tyne	Malpeque Bay	94	0.844	0.847	0.002	15.13	12.100
34	Granville	New London Bay	54	0.829	0.806	0.038	11.88	10.900
35	Tims	Tracadie Bay	69	0.845	0.838	0.016	12.75	11.470
36	Pipers	Tracadie Bay	60	0.838	0.824	0.025	13.00	11.690
37	SWMiramichi	Miramichi	63	0.837	0.864	-0.023	12.63	11.490
38	NWMiramichi	Miramichi	67	0.842	0.834	0.018	13.38	11.700
39	Restigouche	Chaleur Bay	89	0.843	0.831	0.020	13.63	11.390
40	Magdalene	Gulf of St. Lawrence	40	0.814	0.798	0.033	10.75	10.500
41	PtAmal	Port aux Port Peninsula	80	0.824	0.823	0.008	13.00	11.410
42	St. Mary's Harbour	Labrador	55	0.679	0.701	-0.024	8.25	7.630
43	StAnthony	Northern Peninsula	76	0.771	0.723	0.071	9.00	8.210
44	Chuff Brook	Bonavista Bay	93	0.722	0.699	0.037	10.88	9.070
45	Gambo	Bonavista Bay	80	0.738	0.668	0.102	10.88	9.560
46	Traverse	Bonavista Bay	93	0.706	0.668	0.059	10.38	8.970
47	SalmonCove	Conception Bay	96	0.627	0.621	0.016	9.63	8.140
48	Biscay	St. Mary's Bay	96	0.693	0.680	0.024	9.00	7.450
49	Holyrood	Holyrood Pond	96	0.701	0.700	0.008	9.38	8.100
50	Deer	Holyrood Pond	41	0.676	0.695	-0.015	8.00	7.770
51	PathEnd	Holyrood Pond	96	0.705	0.706	0.004	9.50	8.200
52	Salmonier	St. Mary's Bay	93	0.685	0.712	-0.033	9.63	8.200
53	Colinet	St. Mary's Bay	96	0.689	0.712	-0.028	10.13	8.330
54	NHSMB	St. Mary's Bay	94	0.679	0.714	-0.045	9.88	8.320
55	SEPL	Placentia Bay	96	0.646	0.648	0.003	8.88	7.420
56	LongsHr	Placentia Bay	79	0.667	0.689	-0.027	7.13	6.280
57	NHPB	Placentia Bay	80	0.643	0.635	0.019	8.13	7.240
58	SaltPond	Placentia Bay	92	0.753	0.727	0.040	9.38	7.890
59	Garnish	Fortune Bay	74	0.788	0.763	0.039	10.88	9.750
60	Little	Bay d'Espoir	96	0.670	0.685	-0.017	9.38	7.780
61	Conne	Bay d'Espoir	96	0.673	0.657	0.030	10.63	8.150

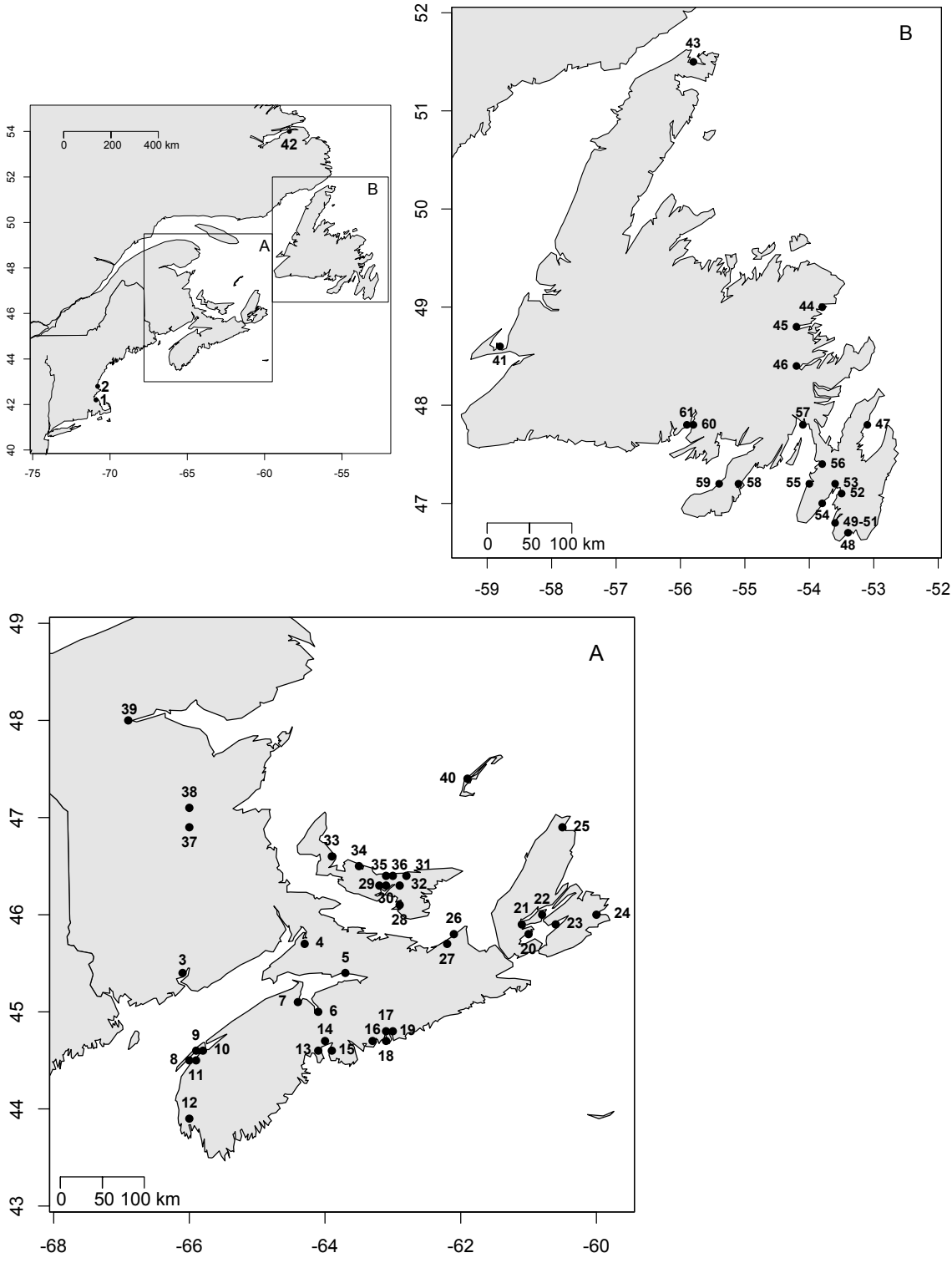

Figure 5.1 Maps of sample locations used in the current microsatellite survey. Numbers refer to sample sites listed in Table 5.1.

5.2.2 DNA Extraction and Microsatellite Genotyping

DNA was extracted using the protocol of Elphinstone *et al.* (2003), modified to work on a 96-well filter plate on a robotic liquid handling system (Perkin Elmer). PCR amplification of nine microsatellite loci (*MGPL-Omo1*, *MGPL-Omo2*, *MGPL-Omo3*, *MGPL-Omo4*, *MGPL-Omo5*, *MGPL-Omo9*, *MGPL-Omo11*, *MGPL-Omo15* and *MGPL-Omo16*) followed the procedures outlined in Coulson *et al.* (2006). Briefly, 5-10 μL volumes were used containing 20-100 ng DNA, 1.5 mM $MgCl_2$, 80 μM of each dNTP, 0.5 U *Taq* DNA polymerase (Applied Biosystems), 0.3 μM of each primer (one primer in each pair labelled with HEX dye) and 1x PCR buffer. PCR products were visualized on 6% denaturing polyacrylamide gels with formamide and imaged on an FMBioII system (Hitachi Genetic Systems). Alleles were scored by eye against a known size ladder with redundant and positive controls to ensure repeatability and accuracy of results.

5.2.3 Genetic Diversity and Measures of Differentiation

The program MICROCHECKER (van Oosterhout *et al.* 2004) was used to test for null alleles, large-allele dropout and scoring errors due to stuttering. Tests for genotypic disequilibrium between all pairs of loci was conducted with the program GENETIX version 4.05 (Belkhir *et al*. 2002) with 1,000 permutations for each locus comparison, both over all populations and for each population individually. The program FSTAT version 2.9.3.2 (Goudet 1995) was used to calculate the number of alleles, observed and expected heterozygosities and departure from Hardy-Weinberg equilibrium, the latter of which was assessed with 10,000 randomizations. Allelic richness was calculated using rarefaction in the program HP-RARE (Kalinowski 2005),

standardized to a common size of 34 individuals (based on the smallest sample size of individuals typed for a locus). Pairwise measures of genetic differentiation (F_{ST} and Jost's *D*) were calculated with GENETIX version 4.05 (Belkhir *et al.* 2002) and the program GENODIVE (Meirmans & Van Tienderen 2004), respectively. In cases of multiple testing, the false discovery rate (Narum 2006) was used to correct initial alpha values ($\alpha = 0.05$).

5.2.4 IDENTIFYING THE NUMBER OF CLUSTERS

Two Bayesian based clustering methods were used to identify genetically similar groups of smelt. BAPS (Corander *et al.* 2003) was used to assess the most likely number of groups (*K*) for the entire data set. Both non-spatial and spatial clustering of groups of individuals (i.e. putative populations) was conducted with the maximum number of assumed populations set iteratively at 1-25, 30, 40, 50, and 61 with three replicates for each *K* and the grouping with the highest calculated likelihood was chosen. After the initial analysis identified the putative number of groups, admixture analysis (Corander & Marttinen 2006) was conducted on the most likely *K* by running 200 iterations to estimate admixture coefficients for the sampled individuals with 200 simulated reference individuals per population and 20 iterations to estimate admixture for reference individuals. Additionally, the program STRUCTURE (Pritchard *et al.* 2000; Falush *et al.* 2003) was also used to estimate the most likely number of groups. For the overall dataset, three replicates for each of $K = 1 – 25$ were initially examined, with a burn-in of 50,000 followed by 200,000 iterations. The number of groups was assessed by the plateau of the ln Pr($X|K$) (Pritchard *et al.* 2000). Both these methods were used as

different clustering methods may yield slightly different results (Latch *et al*. 2006; Francois and Durand, 2010). Clustering was done in a hierarchical fashion until no further groups could be resolved. Graphical results were plotted using the program DISTRUCT (Rosenberg 2004).

5.2.5 Spatial Population Genetic Structure

A principal component analysis (PCA) was conducted on all 61 populations as well as regional groups (see Results), with the program PCA-GEN v.1.2 (Goudet 2005). Analysis of molecular variance (AMOVA) was conducted to compare the partitioning of genetic variance spatially among and within populations groups for the mainland *vs.* Newfoundland, the resolved BAPS clusters and the resolved STRUCTURE clusters. All AMOVAs were executed in Arlequin 3.5.1.2 (Excoffier & Lischer 2010) with 10,000 permutations. To estimate the effects of geographical distance on genetic differentiation, two approaches were taken. First, pairwise measures of linearized F_{ST} [$F_{ST}/(1 - F_{ST})$; Rousset 1997] was regressed against pairwise geographic distances and evaluated in the program IBD version 1.52 (Bohonak 2002) and significance was assessed with Mantel tests (Mantel 1967) using 10,000 permutations of the matrices. Additionally, spatial autocorrelation analysis was conducted in GenAlEx 6.5 (Peakall & Smouse 2006, 2012)) with 999 permutations, 1,000 bootstraps and 50 km distance classes (up to 1500 km). Measures of geographic distance were calculated using Google Earth™ and following within 5 km of the coastline (see Bradbury *et al.* 2008b) between all pairs of populations.

5.3 RESULTS

5.3.1 BROAD-SCALE GENETIC DIVERSITY AND DIFFERENTIATION

Evidence for linkage disequilibrium was detected between *MGPL-Omo3* and *MGPL-Omo16* with all populations pooled ($P < 0.0001$) as well as in greater than 50% of individual population comparisons and remained globally significant after sequential Bonferroni correction. Therefore *MGPL-Omo16* was omitted for the remainder of the analyses, although results were consistent when *MGPL-Omo3* was removed instead. Therefore only eight loci were used for all subsequent analyses.

The microsatellite loci exhibited moderate to high levels of polymorphism. Average number of alleles, allelic richness, population sample size and observed and expected heterozygosity for each sampling location are shown in Table 5.1. The allelic richness, standardized to a common size of 34 individuals, ranged from 6.3 to 12.6 and average observed heterozygosity ranged from 0.621 to 0.875. There were 12 populations that deviated significantly from HWE ($P < 0.05$), however none of these remained statistically significant following correction for multiple tests. Possible null alleles were detected in 16 samples. These were, however, not confined to any particular population and were distributed across most loci. MICROCHECKER did not detect any evidence of large-allele dropout or scoring errors.

Levels of genetic diversity (H_O and A_R) were significantly lower in Newfoundland populations compared to the mainland (1000 randomizations, $P < 0.001$). The exception to this was the Pt. Amal population in southwestern Newfoundland, which showed levels of variability within the range observed throughout the mainland (Figure 5.2). Associated with this difference in diversity was a significantly higher ($P < 0.001$)

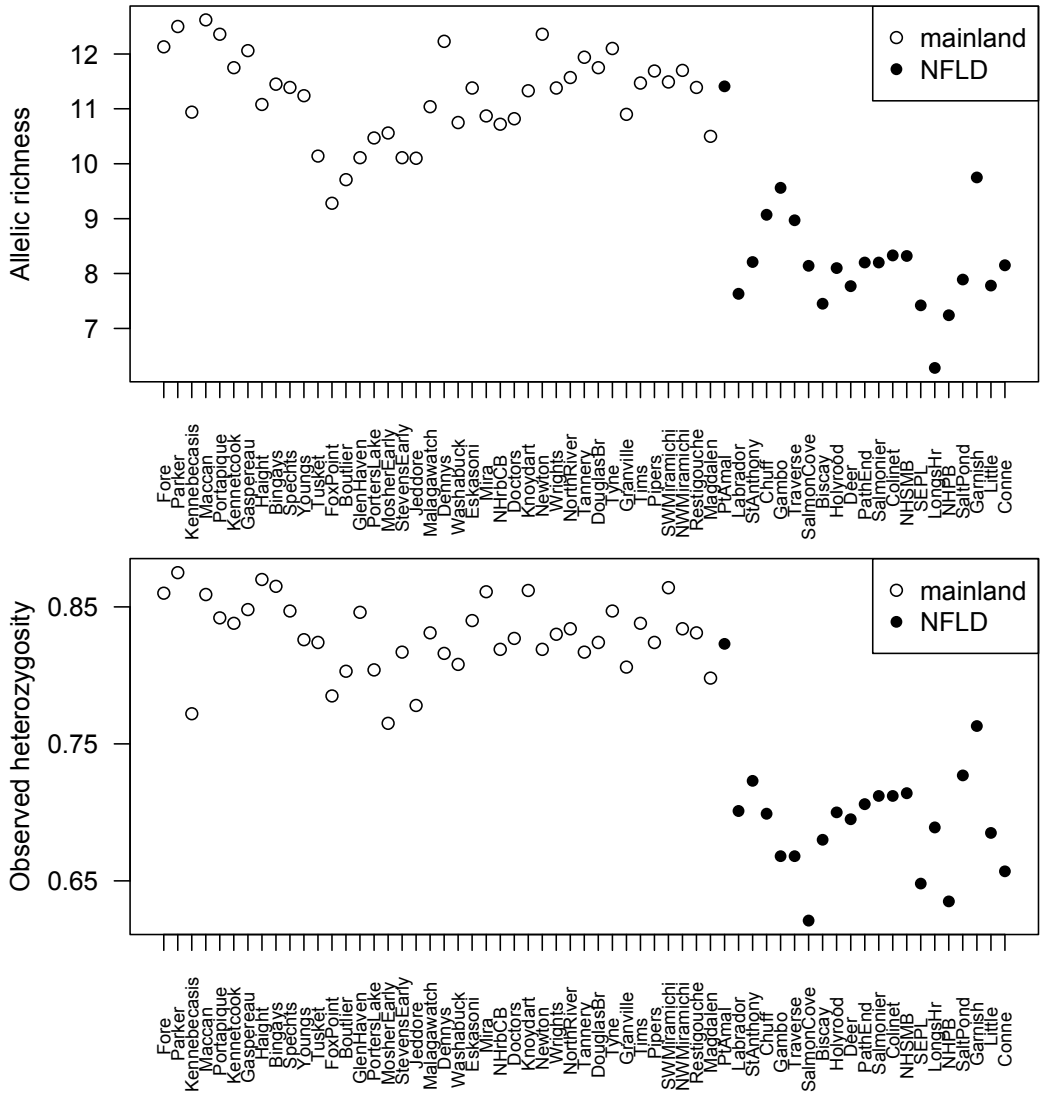

Figure 5.2 Comparison of measure of allelic richness (top) and observed heterozygosity (bottom) among sites.

level of population structure (F_{ST}) within Newfoundland (average pairwise F_{ST} = 0.124, range = 0.001 - 0.232) compared to the mainland (average pairwise F_{ST} = 0.021, range= -0.002 – 0.099) with an average F_{ST} across the entire sampled range of 0.065 (Figure 5.3). Given the observed differences in genetic diversity between these regions, the application of a measure of differentiation that is independent of heterozygosity seems particularly warranted (see Chapter 2). However, even when applying Jost's *D* to these data, the general pattern remained; however, estimates of *D* increased well beyond those of F_{ST} in both regions (Figure 5.3a) and averaged 0.345 (range = 0.003 – 0.636) and 0.108 (range = -0.012 – 0.434) for Newfoundland and the mainland, respectively (Figure 5.3b) with an overall average of 0.236.

5.3.2 IDENTIFYING THE NUMBER OF CLUSTERS

The clustering analysis implemented in BAPS identified 17 distinct genetic clusters (Figure 5.4) (-ln $k(17)$ = 136 792), with a 0.998 probability compared to all other values tested. This result was consistent with our without the use of geographic location to assist in the clustering, and was also consistent with and without the inclusion of Omo11 (see Chapter 2; Bradbury *et al.* 2011). Of the 17 groups identified, 13 were composed only of Newfoundland populations, and were generally associated with estuary or bay-scale structuring. A14th cluster found in Newfoundland was that consisting of Pt. Amal, which grouped with a large cluster of mainland samples ('Gulf'; see below). Within the mainland, there were four distinct groups composed of: (1) the Fore and Parker rivers in Massachusetts along with the Bay of Fundy, St. Mary's Bay and Tusket River [(hereinafter referred to as 'BoF/GoM' (Bay of Fundy/Gulf of Maine)], (2) all sites

sampled along the Atlantic Coast of Nova Scotia (hereinafter referred to as 'Atlantic'), (3) all sites from Cape Breton and Bras d'Or Lakes along with the Gulf of St. Lawrence (hereinafter referred to as 'Gulf') and (4) a group consisting only of the site from the Magdalen Islands. Figure 5.5 shows the average levels of genetic structure within each of the broad regions within the mainland compared with the two broad groups identified in Newfoundland (S1 and S2 from Bradbury *et al.* 2011). Among mainland regions, the GoM/BoF was most distinct and in Newfoundland greater structure was evident among the S2 genetic cluster, which spans a larger geographical scale in Newfoundland.

The STRUCTURE analysis revealed the most likely K =12 (Figure 5.6a & b). This was only marginally higher than K values on either side of it and the bar plot (Figure 5.6a) contained more noise than that observed from the BAPS clustering results. However, within Newfoundland, clustering were generally consistent with bay-scale levels of structuring, as observed with BAPS. There was some indication of possible clustering within the mainland portion of the range, however there appeared to be considerable noise in the signal. A separate STRUCTURE run on just mainland populations revealed the most likely K = 1 (Figure 5.6b). When the Newfoundland sites only were considered in STRUCTURE, the most likely K = 14 (Figure 5.6c), a result consistent with the findings of Bradbury et al. (2008b) despite that study not using Omo11 and consistent with the BAPS clustering within Newfoundland (Figure 5.4).

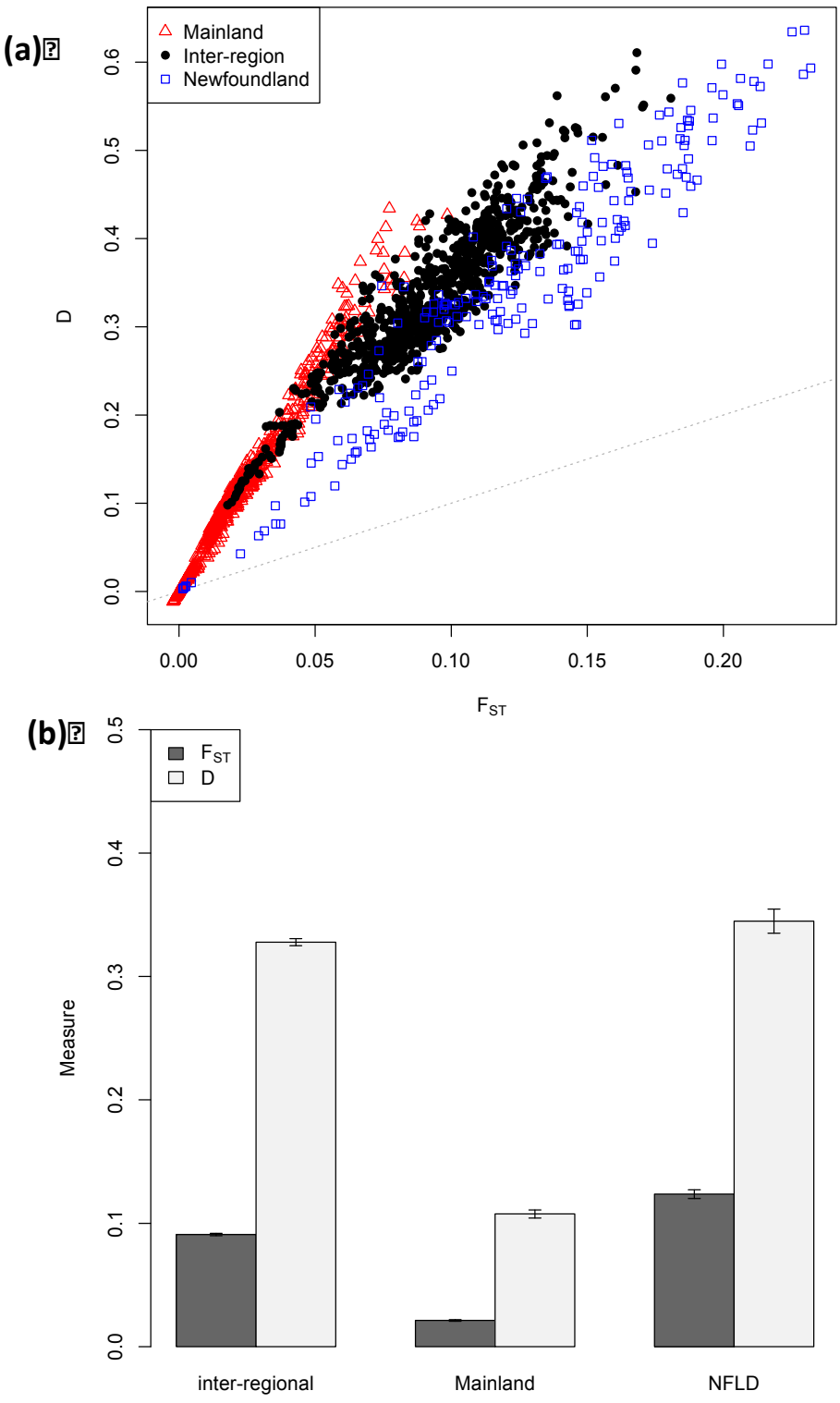

Figure 5.3 (a) Relationship between F_{ST} *vs*. Jost's *D* overall across the dataset and (b) average values within the mainland and Newfoundland.

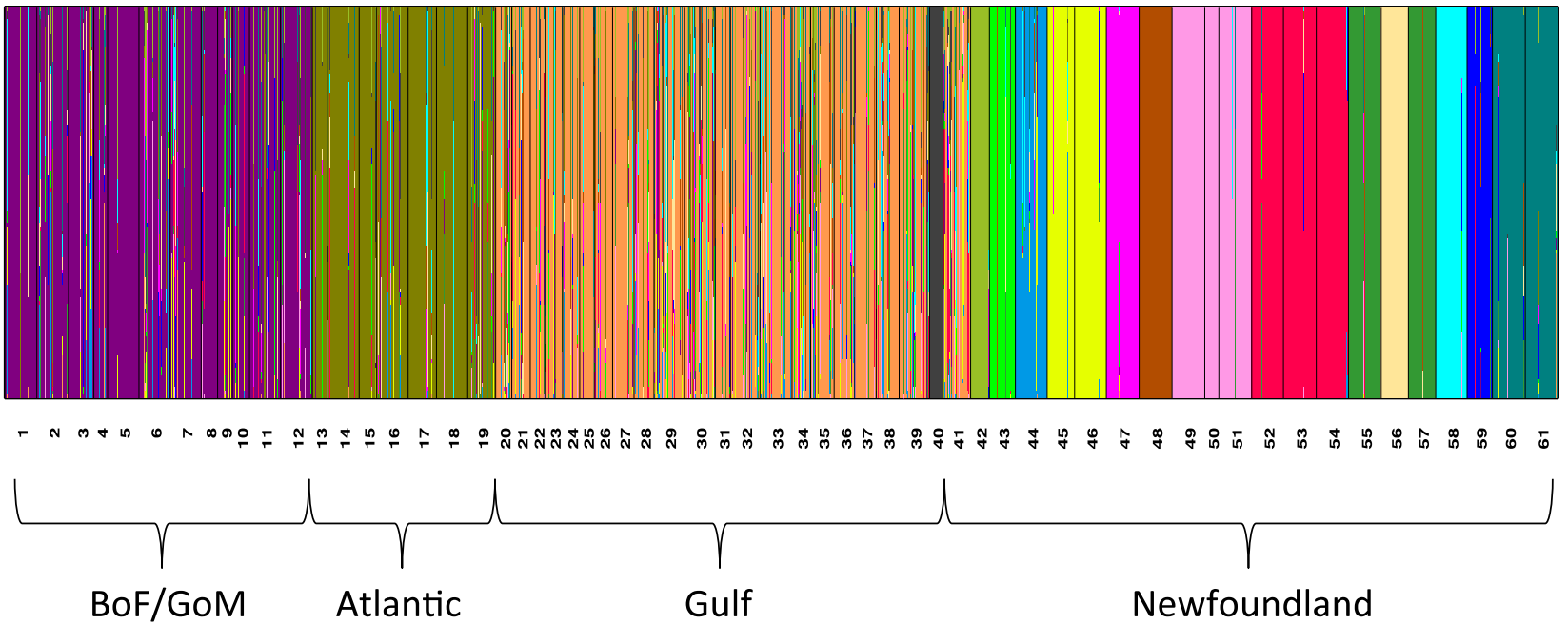

Figure 5.4 Results of the BAPS admixture clustering analysis. The most likely number of groups identified was $K = 17$. A subsequent clustering analysis of the BoF/GoM group indicated separation between two clusters (one composing populations 1-7 and the other 8-12). Clustering of the 'Gulf' group only further resolved a difference between Pt. Amal (site # 41) and the rest of the 'Gulf', while no further separation was possible for the 'Atlantic' group (results not shown).

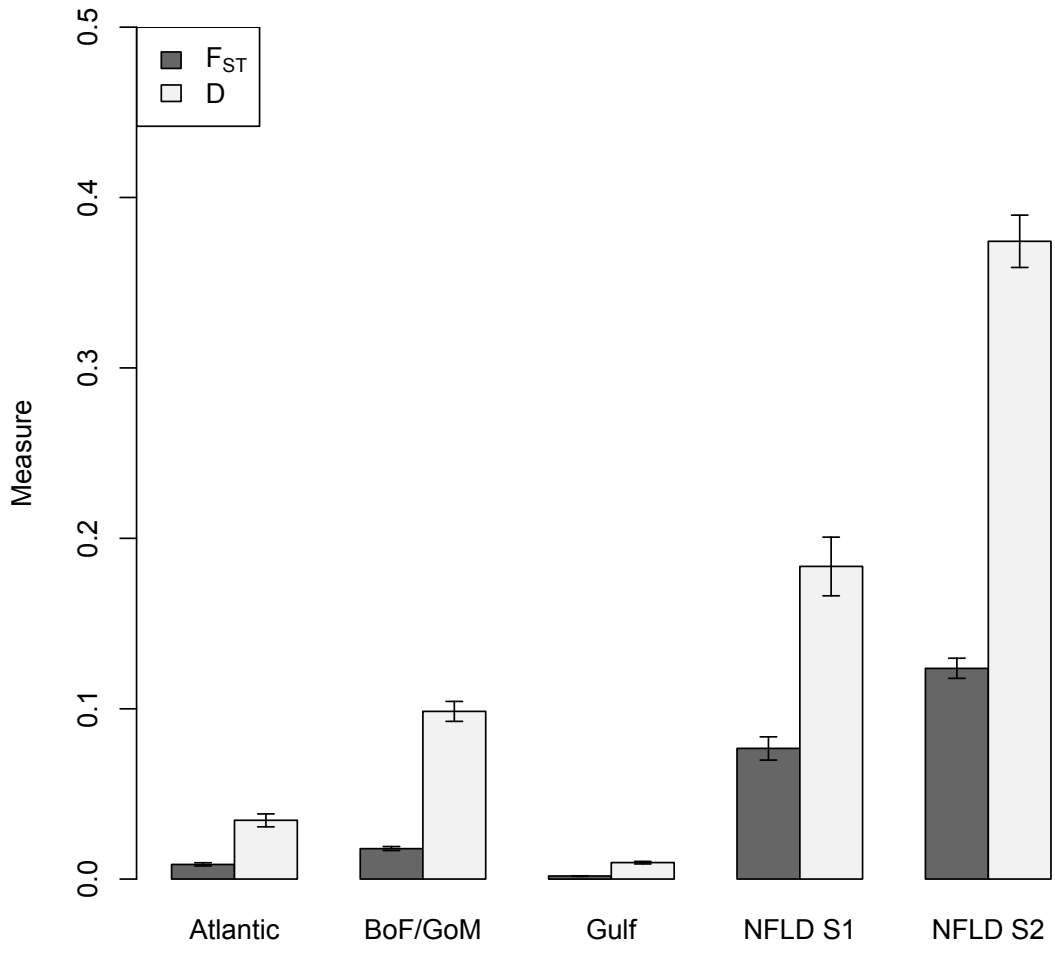

Figure 5.5 Comparison of within group genetic structure (F_{ST} and D) for three mainland regions and two broad Newfoundland genetic clusters.

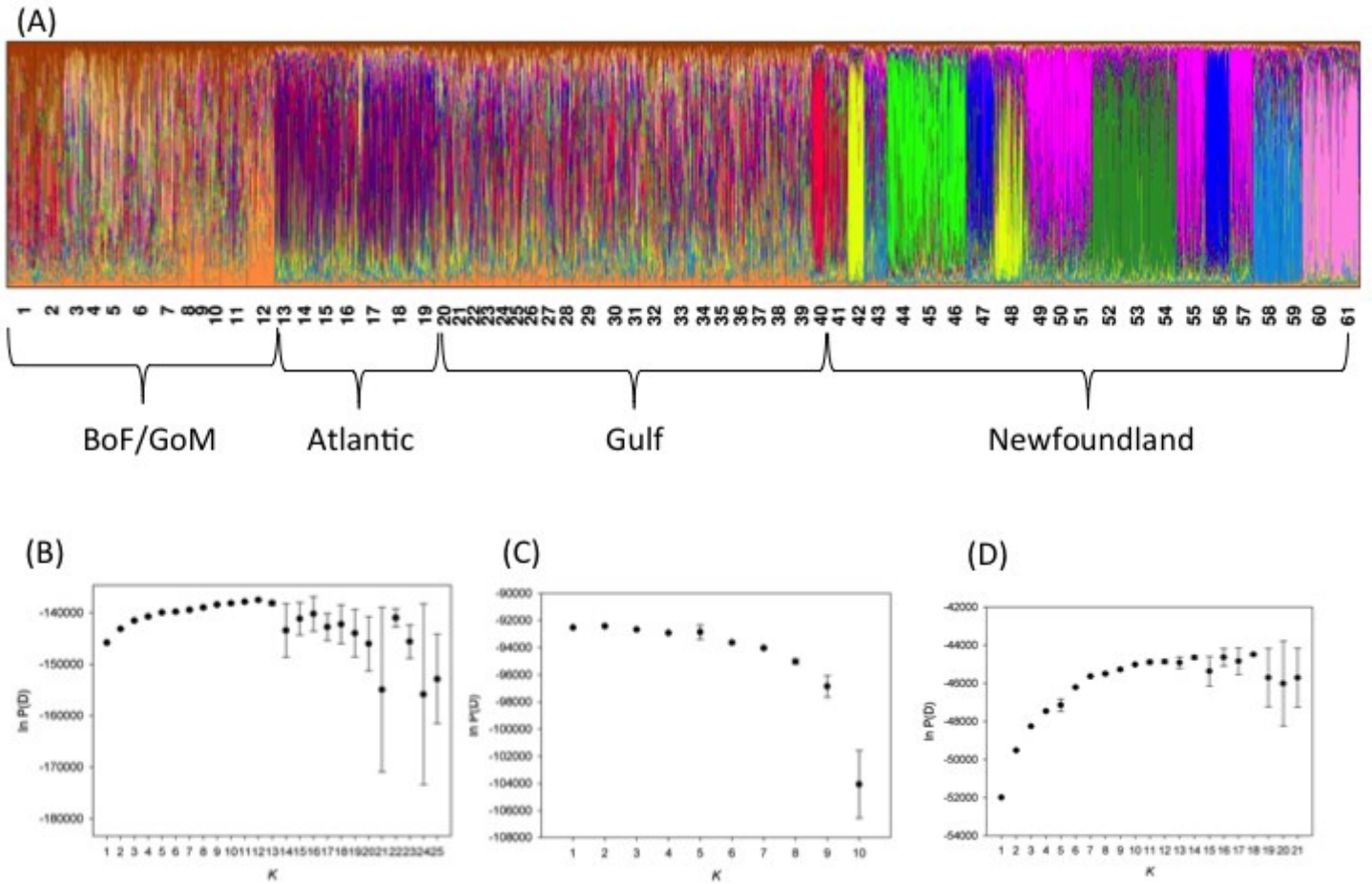

Figure 5.6 Results of STRUCTURE clustering showing (A) $K = 12$ clusters and the likelihood plots for (B) the entire dataset, (C) the mainland sites only (site # 1-40) and (D) Newfoundland only (site # 41-61).

5.3.3 SPATIAL POPULATION GENETIC STRUCTURE

Principal component analysis on the entire dataset primarily isolated the Avalon Peninsula sites in Newfoundland from all remaining sites (Figure 5.7a) (see Bradbury *et al.* 2011; Chapter 3) and this pattern generally held when excluding Omo11 (data not shown). Given the differences in genetic diversity and differentiation between Newfoundland and the mainland, PCAs were subsequently conducted separately on each region. The distinction of the Avalon Peninsula was apparent regardless of the inclusion or not of Omo11 however, there was also further separation of the northeast and south shore sites (Figure 5.7b). Among mainland sites, the PCA resolved grouping consistent with that from the BAPS clustering analysis into three main regions ('GoM/BoF', 'Atlantic', and 'Gulf') plus the isolated Magdalen Islands (Figure 5.7c). As detailed descriptions of population genetic structure in Newfoundland smelt can be found elsewhere (Bradbury *et al.* 2008b, 2011), I focus in more detail on the mainland portion of the species' range.

The average F_{ST}s within the 'Atlantic' and 'GoM/BoF'groups were 0.009 (range = 0.00 – 0.016) and 0.018 (range = 0.0004 – 0.036), respectively while within the 'Gulf' F_{ST} average 0.002 (-0.0022 – 0.009) (Figure 5.7). The Magdalen Island location generally displayed the largest F_{ST}s to all other locations within the mainland and excluding this site from the 'Gulf' region reduced the average within-Gulf F_{ST} to 0.007 (-0.0022 – 0.062). Due to the large range encompassed by these regions (all contained some non-significant F_{ST}) permutation tests did not reveal significant intra-regional differences in F_{ST} (10 000 permutations; $p = 0.20$). On the other hand, permutation tests for differences in observed heterozygosity and average allelic richness were significantly different

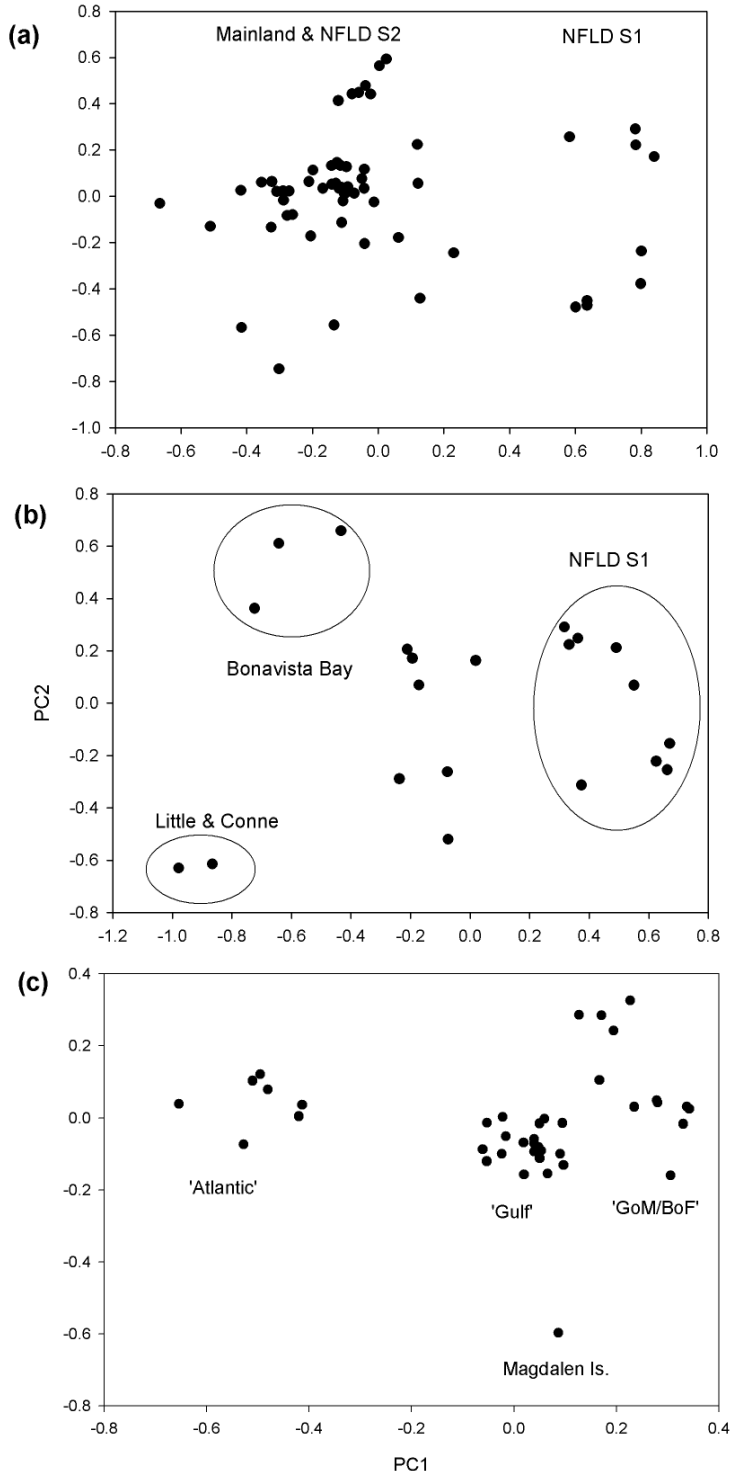

Figure 5.7 Principal component analysis of (a) all samples, (b) Newfoundland samples only (see Bradbury *et al.* 2011 for distinction of S1 & S2) and (c) mainland samples.

across the three groups (both $p < 0.005$; 10,000 permutations) and was driven by the lower levels of diversity for both parameters in the 'Atlantic' group (Tukey's test: $p < 0.001$ for comparisons involving the Atlantic) as there is no difference in these parameters between BoF/GoM and Gulf (Tukey's $p = 0.84$). AMOVAs were conducted for a variety of grouping scenarios and are shown in Table 5.2. Both a scenario in which 14 Newfoundland clusters were considered with or without structuring in the mainland (i.e. the 3 regional groups) explained more of the among group variation than either a simple Newfoundland-mainland comparison or one based on the overall PCA (Figure 5.7a). Additionally the overall PCA, which largely reflected the southern Avalon versus the remaining samples, explained more of the variation than a simple mainland-Newfoundland grouping despite the large differences in genetic diversity among the latter regions (Table 5.2).

There was significant support for an overall isolation-by-distance, however it only explained 14% of the variance, as there was a considerable degree of scatter. Within this pattern, however, IBD differences between the mainland and Newfoundland were apparent (Figure 5.8a), with the latter having previously been documented (Bradbury *et al.* 2006, 2008b). Within the mainland, an overall plot of isolation by distance was not statistically significant (Figure 5.8b; Mantel test: $r = 0.11$, $P = 0.09$), despite a large proportion of significant pairwise comparisons (624 out of 780; of which 566 remained significant after correction for multiple tests). In light of the three regions identified within the mainland, however, effects of IBD were assessed for each and revealed different patterns among regions (Figure 5.8c). It should be noted that due to the higher differentiation of the Magdalen Islands from the rest of the 'Gulf' and its relative

Table 5.2 Analysis of molecular variance (AMOVA) results comparing genetic variation among groups of smelt according to different clustering of groups

Grouping	Source of variation	d.f.	Percent variation	Fixation index	*P*-value
Mainland-Newfoundland	Among groups	1	2.09	0.0805	<0.0001
	Among populations within groups	59	5.96	0.0609	<0.0001
	Within populations	8313	91.95	0.0209	<0.0001
Overall PCA	Among groups	1	5.4	0.1056	<0.0001
	Among populations within groups	59	5.2	0.0551	<0.0001
	Within populations	8313	89.4	0.0534	<0.0001
BAPS (17 groups)	Among groups	16	6.97	0.0798	<0.0001
	Among populations within groups	44	1.01	0.0108	<0.0001
	Within populations	8313	92.02	0.0697	<0.0001
STRUCTURE (15 groups)	Among groups	12	7.6	0.0986	<0.0001
	Among populations within groups	48	2.3	0.0242	<0.0001
	Within populations	8313	90.1	0.0762	<0.0001

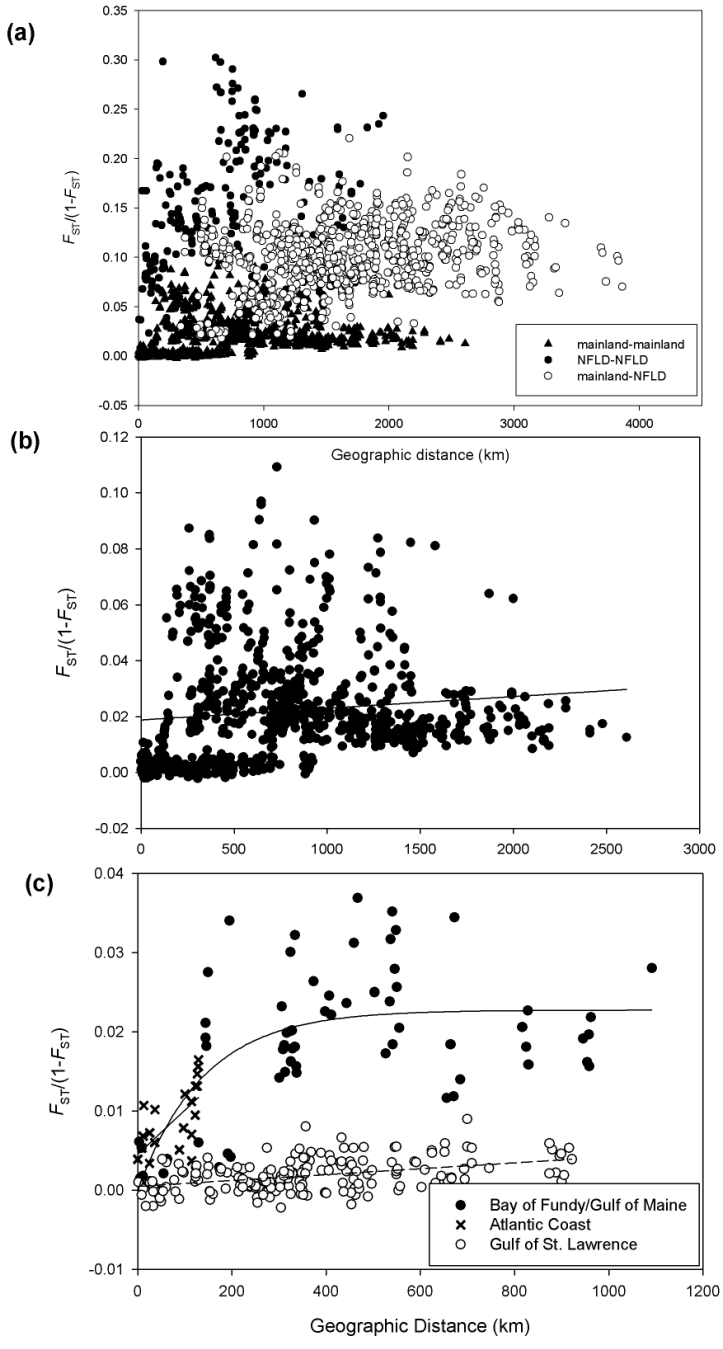

Figure 5.8 Isolation by distance plots for (a) the entire data set, (b) mainland only sites overall and (c) within the three mainland regions identified from the clustering analysis. See Bradbury *et al.* (2008b) for detailed description of IBD in Newfoundland.

geographical isolation, this site was excluded for this purpose. While the spatial scale is quite limited for the 'Atlantic' group, the BoF/GoM and the 'Gulf' region demonstrated very different patterns. There was a significant and positive slope of the IBD for the 'Gulf' samples (Mantel test: $r = 0.42$ $p < 0.0001$) across the entire region, in contrast to the BoF/GoM that showed a non-linear trend in the IBD with a plateau around 200 km, which explained more of the variance than a linear model (Mantel test linear: $r = 0.46$, $p < 0.0091$ vs. nonlinear model: $r = 0.69$, $p < 0.0001$). A significant IBD was also apparent in the Atlantic group (Mantel test: $r = 0.61$, $p < 0.0001$) however the geographical range of samples only encompassed 200 km. Spatial autocorrelation analysis failed to detect significant autocorrelation from small-scales (~50 km) to across the entire mainland portion of the range, despite distance classes approaching significance at several hundred kilometers. In Newfoundland, Bradbury *et al.* (2008b) detected significant spatial autocorrelation and isolation-by-distance up to ~200 km, suggesting a limit to gene flow beyond this scale (Figure 5.6c, reproduced from Bradbury *et al.* 2008b).

5.4 DISCUSSION

Defining the spatial scale at which species are structured is a fundamental aspect of molecular ecology and can have important implications for management. While a growing body of work demonstrates many marine species are geographically structured (e.g, Nielsen *et al.* 2003; Ruzzante *et al.* 2006; Bradbury *et al.* 2010; Hauser & Carvalho 2008), more needs to be done to understand the relative contribution of underlying mechanisms, both historical and contemporary. The absence of physical barriers, in some settings and associated life-history stages that should facilitate dispersal (i.e. pelagic

larvae) are clearly not the only determinant characteristics regulating connectivity between demes. In particular, the relative influences of gene flow and natural selection at varying spatial scales are still largely unresolved.

Clear genetic differences exist among regional groups of rainbow smelt. This is most obvious in the differences in genetic diversity and genetic differentiation between Newfoundland and the mainland. Regardless of the exact number of population clusters within Newfoundland (~12-14) the genetic structure was largely consistent with differences among bays or estuaries, which has been reported previously and corroborated with otolith elemental analysis (Bradbury *et al*. 2008b, c). In addition to the differences in diversity, population structure within Newfoundland also appears largely driven by the occurrence of two predominant races in a zone of secondary contact, and demonstrates the uniqueness of populations occurring on the Avalon Peninsula. However, even against this historical backdrop of genetic differentiation the differences among bays within Newfoundland explained more of the variance than the phylogeographic lineages (see AMOVA results).

On the other hand, population structure within the mainland was significantly weaker than in Newfoundland, an observation that has been reported elsewhere (Bradbury *et al*. 2006a, Kovach *et al.* 2013). Despite this weaker population structure, there was still evidence for broad, regional differences within the mainland. While the STRUCTURE analysis failed to resolve more than a single group, the clustering with BAPS resolved three main groups (and a fourth composed solely of the Magdalene Islands). While differences in clustering algorithms are to be expected (Latch *et al.* 2006), the resolution of four groups within the mainland appears to be supported by the

PCA analysis of this region (Figure 5.7c). In addition, differences in patterns of IBD among these regions further support regional population structure within the mainland. Indeed, Kovach *et al.* (2013) resolved four groups (with BAPS and STRUCTURE analysis) within their study area, which was more geographically restricted than the present study. Furthermore, Kalinowski (2011) demonstrated that STRUCTURE, specifically, can fail to resolve genetic clusters depending on the degree of divergence and due to differences in sample sizes among clusters. While the inclusion of the mainland as a single or 3 distinct groups had a minimal impact on the amount of variation explained by the AMOVAs, this could largely be due to the number and divergence of the different Newfoundland groups in contributing to the among group variation.

Overall, the results of the genetic survey of rainbow smelt presented here demonstrate how varied patterns of genetic differentiation can be at the intraspecific level. Differences in the scale of dispersal, past historical events, contemporary landscape features, oceanographic conditions and adaptive variation, amongst other factors can interact to give rise to complex patterns of population structure. I consider, in turn, both the historical phylogeographic and contemporary signatures contributing to the overall patterns and diversity throughout a species range.

5.4.1 The Role of History and Retention of Past Genetic Signatures

Clearly, historical events have left their legacy in the patterns and diversity of structuring observed across the range of rainbow smelt. The degree of genetic structuring observed in Newfoundland appears to be quite high for this, and related species (Kovach

et al. 2013, McLean & Taylor 2001; Beacham *et al.* 2005) and influenced by the recolonization of the island by two glacial races, which appear to show signs of differential introgression highlighted by some degree of cyto-nuclear discordance (Chapter 3; Bradbury *et al.* 2011; Toews & Brelsford 2012). However even among populations within a race, the levels of genetic differentiation are still relatively high compared to the rest of the species range. In a closely related species, capelin (*Mallotus villosus*), Colbeck *et al.* (2011) found evidence for historical introgression based on nuclear AFLP markers among previously identified mtDNA clades of while others showed a signature of divergence without gene flow. Similar patterns of high levels of genetic structuring in Newfoundland have been observed in other species, such as Atlantic salmon (Palstra *et al.* 2007), which have also been documented to be composed of distinct glacial relicts (Verspoor 2005).

Even within the mainland portion of the species range considered here, where differentiation is much weaker, and a single race predominates, there appears to be a legacy of historical events. The closer relationship between the GoM/BoF and Gulf groups may represent previous contact among these regions. There is some evidence (Shaw *et al.* 2006) that the isthmus connecting New Brunswick and Nova Scotia had previously been submerged, thereby creating a corridor for connection and dispersal between the Bay of Fundy and Gulf of St. Lawrence. Such a connection is also largely supported by the mtDNA evidence (Chapter 3), whereby both these regions had contribution from both ‘A’ and ‘B’ mtDNA types. On the other hand, the ‘Atlantic’ group was virtually fixed for the ‘A’ mtDNA type with the exception of a low

contribution (~10%) from the 'B' type within a single estuary and limited to a distinct temporal run of smelt.

The legacy of historical vicariance and previous glacial cycles have increasingly been implicated in surveys of nuclear variation (e.g. Rey & Turgeon 2007; Duvernell *et al.* 2008; McCusker & Bentzen 2010), which have typically been used as an alternative to mtDNA to highlight contemporary patterns of population structure. A recent study by Hasselman *et al.* (2013) resolved a clear genetic break between northern and southern populations of American shad (*Alosa sapidissima*) that was closely associated with the maximal extent of ice during the last glaciation. However, this phylogeographic break is also confounded with a boundary to the south that is concordant with different life history strategies (i.e. semelparity vs. iteroparity), highlighting the difficulties involved with disentangling competing factors underlying different population genetic signatures.

5.4.2 Contemporary Factors Structuring Populations

Smelt population structure has previously been linked to local retention areas, defined as the scale at which an individual completes its life-cycle, by other authors based upon morphology (Frechet *et al.* 1983a), parasites (Frechet *et al.* 1983b), genetic markers (Baby *et al.* 1991; Bernatchez and Martin 1996; Bradbury *et al.* 2008b) and otolith elemental analysis (Bradbury *et al.* 2008c). Given the diversity in the size and shape of retention areas throughout the range of smelt, this would be expected to translate into different scales of connectivity. The large tidal flushing within the Bay of Fundy likely results in strong selection pressures for retentive processes. Indeed behavioural modification, such as vertical migration in smelt (e.g. Bradbury *et al.* 2006b) has been

implicated as a mechanism for larval retention (Ouellet and Dodson 1985; Laprise and Dodson 1989). Similarly, the counter-clockwise oceanographic circulation patterns within the Gulf of St. Lawrence and the relatively continuous coastline of the area likely with abundant habitat for smelt could result in the observation of large-scale dispersal in this retention zone. The separation of the 'Atlantic' group from the Gulf to the north is likely influenced by the flushing through the Cabot Strait and to the south, the meeting of the Grand and Georges Bank at the outflow to the Gulf of Maine. Given the relatively linear coastline of Nova Scotia along the Atlantic Coast, more samples of smelt from this area spanning >200 km sampled here would be of interest. Many of the smelt populations here are likely contained within small bays or inlets and would lend to a further assessment of retentive processes acting along this region.

Kovach *et al.* 2013 recently reported on a genetic analysis of rainbow smelt from 18 populations throughout the coastal area of Maine and Massachusetts. They found levels of genetic differentiation comparable to those seen among the mainland samples considered here. Furthermore, they inferred four genetic groups, based on Bayesian clustering analysis, and attributed these groups to hydrographic conditions generating areas of larval retention. They found a significant effect of IBD and spatial autocorrelation, indicating dispersal at scales of up to 200 km. Interestingly, this is approximately equivalent to the scale of the plateau in the IBD for the Bay of Fundy/Gulf of Maine reported here and for Newfoundland smelt according to Bradbury *et al.* (2008), despite the larger genetic differences in Newfoundland compared to the mainland portion of the range explored here and in Kovach *et al.* (2013). Finally, similar retentive processes have been invoked to explain genetic differentiation within the St. Lawrence

estuary (e.g. Bernatchez & Martin 1996), however, like the situation in Newfoundland, this is a zone of secondary contact among mtDNA types on opposite sides of the St. Lawrence River.

While the identification of isolation by distance should seem straightforward, other factors can influence the ability for such signals to be resolved. The different scales and shapes of the IBD between the three mainland regions resulted in a non-significant overall relationship when not taking local differences into account. Similarly, the strong IBD relationship in Newfoundland smelt coupled with the weaker relationships among mainland smelt resulted in a significant, but weak range-wide effect of IBD. Therefore combining information from different areas with different underlying patterns may obscure certain signals of population differentiation. Indeed, Kittlein & Gaggiotti (2008) demonstrated that interactions between environmental factors can mask isolation by distance patterns. This demonstrates the usefulness of alternative approaches, such as the Bayesian clustering used here, to highlight regional differences that can then be used to form the basis of broad-scale studies encompassing varied habitats and environments. More recently, Puebla *et al.* 2012, demonstrated that non-significant but positive isolation by distance can be associated with limited dispersal, based on tagging studies of coral reef fishes and suggested when mean dispersal distance is smaller than the potential for dispersal such patterns are compatible with low genetic structure on large geographical scales (Lotterhos 2012).

Clearly to understanding the genetic patterns connecting different populations of a species requires consideration of both historical and contemporary factors. While disentangling the relative roles different factors may play in promoting or restricting

population connectivity, the increasing rate at which genetic information is made available, even for non-model species, will help to shed light into current questions and open up new avenues that have not yet been possible. Furthermore, understanding the population structure of a species requires comprehensive coverage across the entire range and local sampling may fail to capture the true diversity and complexity of processes involved in population differentiation.

CHAPTER 6

TEMPORAL GENETIC DIFFERENTIATION: CONTINUOUS *VS.* DISCONTINUOUS SPAWNING RUNS IN ANADROMOUS RAINBOW SMELT *OSMERUS MORDAX*

6.1 INTRODUCTION

Spatial separation of spawning aggregations has long been recognized as a major determinant of population genetic differentiation. For fishes that occupy continuous or quasi-continuous environments, genetic differences often accrue as a direct function of the spatial distance that separates spawning aggregations, a pattern known as isolation by distance (IBD) (Palumbi, 2003). In theory, however, differences in the timing of reproduction may limit gene flow in a manner analogous to the way in which spatial separation can produce IBD. Hendry & Day (2005) used the term 'isolation by time' (IBT) to emphasize this parallel, and they provided a theoretical framework to describe the population genetic consequences of temporal separation of reproduction. Their theory posits that if the timing of reproduction is heritable, then genetic differences within the population should increase with increasing differences in reproductive time.

The fact that the majority of evidence for reproductive timing as an important determinant of genetic differentiation in fishes comes from Pacific salmon is likely because this group has been so extensively studied and has been found to exhibit considerable heritable variation in reproductive timing within species (reviewed in Hendry & Day, 2005). Genetic differences associated with spawning time have been observed in both hatchery environments (e.g., Fishback *et al.*, 2000) and natural populations (e.g., Woody *et al.*, 2000). Thus far, little is known for non-salmonid fishes

about the heritability of the timing of reproduction, or of the importance of variation in reproductive timing as a determinant of genetic differentiation.

The rainbow smelt (*Osmerus mordax* Mitchill) is a euryhaline species that occurs along the east coast of North America from New Jersey north to Labrador (Scott & Scott, 1988). Most smelt are anadromous; spawning occurs at night in streams and rivers a short distance above the head of tide. Smelt populations inhabit diverse environments and commonly appear as differentiated ecotypes. The most fundamental ecotypic divergence is between anadromous and freshwater-resident forms, but there is also phenotypic variation within each of these life history types (e.g., Saint-Laurent *et al.*, 2003; Lecomte & Dodson, 2004; Bradbury *et al*. 2006).

Although commonly documented smelt phenotypic variation involves morphological traits, variation in the timing of reproduction is also known to occur both among and within populations (McKenzie 1964). Spawning can occur as early as March, and as late as early July in different populations in eastern Canada (Scott & Scott, 1988; personal observations). Within the Miramichi River, McKenzie (1964) documented 'early', 'middle' and 'late' spawning runs that occurred over a two-month period from April to June. Recent observations have revealed smelt spawning up to two months apart within several small coastal streams in eastern Canada (M. Coulson and I. Bradbury, unpublished data), and it seems likely that protracted or bimodal spawning is common in smelt populations.

6.2 MATERIALS AND METHODS

There have been no direct estimates of the heritability of spawning timing in smelt, but stock transfer studies by Rupp & Redmont (1966) suggested that spawning time was under some degree of genetic control. If the timing of spawning in smelt is significantly heritable, then reproductive timing may contribute to the reproductive isolation of populations, and lead to genetic differentiation in a manner similar to that seen in Pacific salmon. To test this prediction, we used two approaches: a simulation and a genetic analysis of two streams that both harbour aggregations of smelt spawning a month or more apart. In one stream, Mosher's Brook in Nova Scotia, there are two distinct spawning runs of smelt: an 'early' run during the second half of April and a 'late' run in mid to late June, with an interval of ~1 month during which little or no spawning is evident. In the second stream, Smelt Brook in Southeast Placentia, Newfoundland, early and late spawning aggregations of smelt also occur at least a month apart; however spawning is continuous over this time interval. In both streams, the early and late spawning components of the population appear to spawn in exactly the same habitat. These two situations suggest different scenarios for genetic differentiation. In Smelt Brook, on-going spawning from mid-April to the end of May suggests a situation analogous to the 'stepping stone' model of population structure, in which temporal 'straying' or 'migration' occurs primarily between closely sequential spawning times. In this scenario, extensive genetic mixing between similar spawning dates might lead to relatively high overall gene flow between the extreme 'early' and 'late' spawning aggregations of smelt. Consequently, genetic differentiation between the early and late extremes of the spawning run might be very weak or undetectable. In Mosher's Brook,

on the other hand, with no spawning occurring for approximately a month, the net gene flow between early and late spawning fish might be lower than in Smelt Brook, with consequently greater genetic differentiation, despite a similar difference in spawning timing, and assuming the same heritability of reproductive timing. These predictions were tested by using microsatellite analysis to compare genetic differentiation between early and late spawning aggregations in the two streams.

The simulation model was constructed to account for two key variables, the difference in timing between 'early' and 'late' spawning smelt, and the number of nights over which individual smelt spawn. The purpose of the model was to determine if significant genetic differentiation between 'early' and 'late' spawning smelt could occur given various biologically plausible values for the key variables, and to examine the relative contribution of model parameters to genetic differentiation. We hypothesized that the number of nights across which individual smelt spawn is a more important factor contributing to genetic continuity within a run than the number of days separating spawning individuals. We used a one-dimensional stepping-stone model analogous to a model of isolation by distance by Palumbi (2003) with spawning events equally spaced over a 30 day period. This model assumed a constant population size (N=1000) with equal gametic contributions per individual and the simulation was run for 10,000 generations to ensure equilibrium conditions were established. Genetic drift and gene flow occurred following Kimura (1980) and the probability of temporal dispersal from night x to y was approximated as $p(x,y,) = \alpha/2 \exp(-\alpha \mid x\text{-}y\mid)$, where $1/\alpha$ is the mean number of nights spawning and varied from one to five nights, following the dispersal kernal used in spatial models (Botsford *et al.*, 2001). Night specific allele frequencies

were estimated and F_{ST} was calculated based on 10 nights of equal spacing, beginning at night one along the stepping-stone, and was averaged over five loci (the single locus model was run five times to simulate five loci).

For the genetic analysis, smelt were sampled from Smelt Brook in Southeast Placentia Bay, Newfoundland and Mosher's Brook in Musquodoboit Harbour, Nova Scotia. Both brooks are small, ~1.5m wide at the mouth and discharge directly into the estuary. Neither brook has any tributaries, and all smelt spawning appears to occur over a ~50m stretch immediately above the upper tide limit. Both brooks are shallow (max. depth < 0.5 m) and only accessible to smelt near high tide. The Smelt Brook samples were collected on 18 April (n = 46) and 23 May (n=52), 2004. The Mosher's Brook 'early' run (n = 59) was collected 21-27 April and the 'late' run (n = 101) was sampled 23 June 2004. All fish were sampled by dip-netting. Individuals were measured, sexed, scale samples taken and fin clipped in the field and tissue was placed directly in 95% ethanol. DNA extraction, PCR and gel electrophoresis conditions for nine smelt-specific microsatellite loci were performed as described in Coulson *et al*. (2006).

For each sampling period, the number of alleles per locus, allelic richness, observed heterozygosity and unbiased estimates of heterozygosity were calculated using FSTAT (v. 2.9.3.2; Goudet, 2002). Samples were tested for conformity to Hardy-Weinberg equilibrium and genotypic linkage disequilibrium as implemented in FSTAT (Goudet, 2002). Genic tests of differentiation among sampling periods were performed using GENEPOP (Raymond & Rousset, 2003). Additionally, genetic divergence between pairwise samples was measured with F_{ST} and level of significance was estimated using a permutation test (1000 permutations) as implemented in Genetix (v. 4.05.2, Belkhir *et*

al., 2000). As further tests for the ability to distinguish between each of the two sampling periods for each location, separate assignment tests were carried out for each locality using GENECLASS2 (Piry *et al.*, 2004) and a factorial correspondence analysis on the multilocus genotype data was performed on all four samples as implemented in Genetix (Belkhir *et al.*, 2000).

6.3 RESULTS

The simulation suggested that significant genetic differentiation could occur between early and late spawning smelt in a single stream, and that the duration of spawning by individual smelt is a more important factor than the amount of time separating the early and late spawning components. In simulation trials, significant genetic differentiation only occurred between smelt aggregations spawning as much as a month apart if individual smelt spawned on three nights or less. Further, if individuals spawned only on a single night, the simulation resulted in an F_{ST} of ~0.01 between groups of smelt that spawn 30 days apart, and yielded detectable differentiation (F_{ST} = 0.005) between groups differing in spawning time by only two weeks. At the other extreme, simulations in which individual smelt spawned over four nights, on average, resulted in zero genetic differentiation between groups of smelt spawning 30 days apart (Figure 6.1).

Microsatellite data collected from individuals at both sites were checked for presence of null alleles and scoring errors using MICRO-CHECKER (van Oosterhout *et al*., 2004). Tests of departure from Hardy-Weinberg equilibrium (HWE) revealed a marginal probability of departure from HWE for Mosher's Brook 'early' at locus Omo3

(*P*-value = 0.0014; tablewide Bonferroni corrected *P*-value of 0.00139). This was due to a larger number of homozygotes than expected and MICRO-CHECKER analyses suggested this could be due to the presence of null alleles. Additionally, tests of linkage disequilibrium revealed a marginally significant association between Omo3 and Omo16, but only for the two Mosher's Brook runs ($P = 0.00035$; Bonferroni corrected *P*-value = 0.000347). All analyses described below were completed both with and without either Omo3 or Omo16 and revealed no changes in levels of significance or outcomes of the analyses, therefore both loci were included in all analyses. All loci were polymorphic across the four populations with 12–22 alleles per locus. The Mosher's Brook runs were more variable than the Smelt Brook run, both in terms of number of alleles (6-20 vs. 3-15) and gene diversity (H_E) (0.716-0.929 vs. 0.244-0.877). This difference in diversity between the Nova Scotia and Newfoundland samples is also evidenced by Bradbury *et al*. (2006).

Pairwise calculations of F_{ST} between 'early' and 'late' samples revealed different results in the two streams. F_{ST} between 'early' and 'late' in Mosher's Brook was low but significantly greater than zero ($F_{ST} = 0.0064$; $P = 0.001$); whereas F_{ST} between 'early' and 'late' in Smelt Brook was non-significant ($F_{ST} = 0.0027$; $P = 0.174$). For comparison, F_{ST} between the two streams was much greater ($F_{ST} = 0.102$-0.117, both $P<0.0001$). Additional tests of population differentiation using genic tests of allele frequency heterogeneity and assignment tests were both consistent with the F_{ST} results. In genic tests, none of the nine locus comparisons between Smelt Brook 'early' and Smelt Brook 'late' were significant, whereas three of the nine locus comparisons were statistically

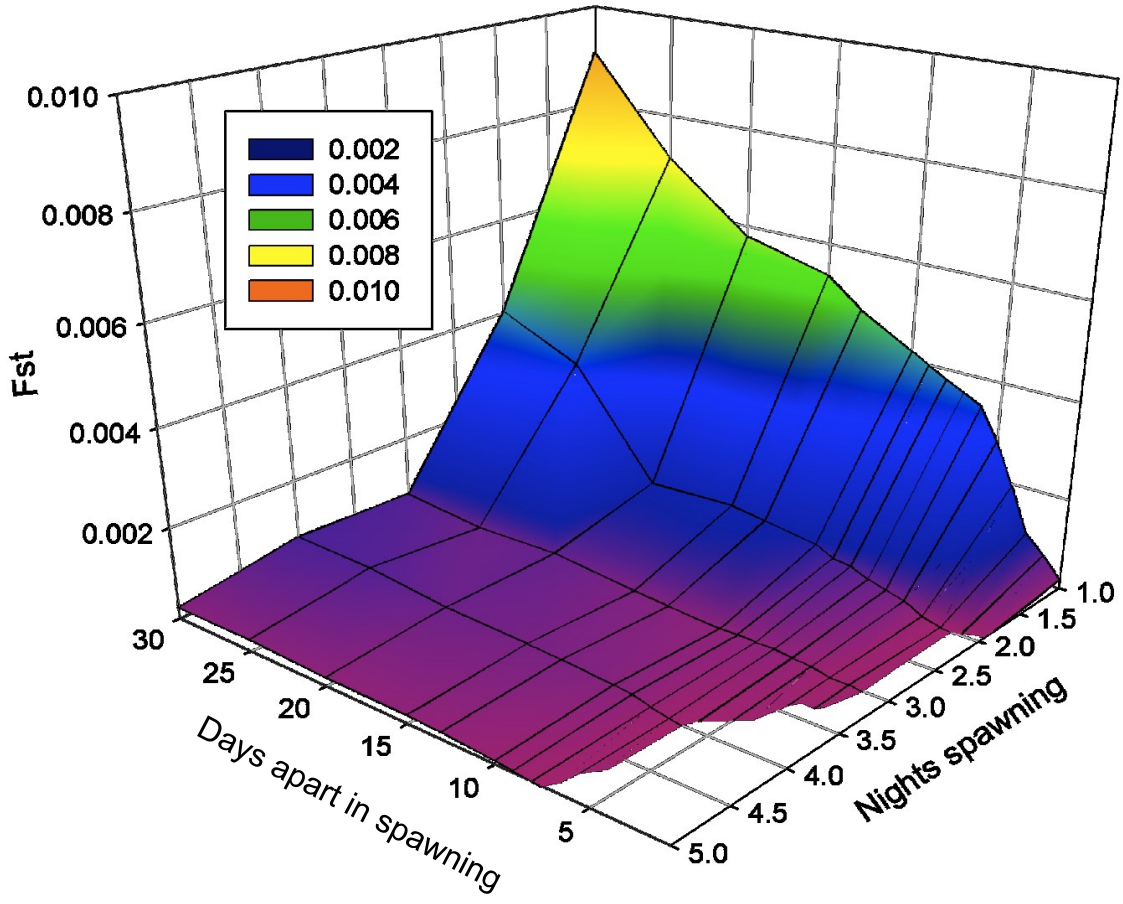

Figure 6.1 Results of the simulation of the effect of the number of consecutive nights spawning and the time on estimates of genetic differentiation.

significant between Mosher's 'early' and 'late' runs. Similarly, assignment tests conducted on the Smelt Brook samples yielded a correct assignment (back to time of spawning) of 54.2% (52 out of 96 individuals) which is not significantly different from random expectations for two populations ($\chi^2 = 0.677$, df = 1, $P = 0.411$). On the other hand, 75.6% (121 out of 160 individuals) of Mosher's Brook fish were correctly assigned to their time of capture, a result that is significantly different from random expectations ($\chi^2 = 37.2$, df = 1, $P < 0.001$).

Finally, a factorial correspondence analysis (FCA) revealed a striking difference in the clustering of the multilocus genotypes of the Mosher's Brook fish compared to those of the Smelt Brook fish (Figure 2). The first axis explained 72.6% of the variation and distinguished the two geographic regions, Newfoundland and Nova Scotia, while the second axis (19.1%) demonstrated a much larger separation between the two Mosher's Brook run times than was evident for the early and late sampling periods from Smelt Brook, which revealed only a weakly suggested separation, primarily along axis 3 (8.3%). The large separation between Newfoundland and Nova Scotia (axis 1) is not surprising given that an analysis of 20 populations spanning Atlantic Canada identified a major break between Newfoundland and the mainland (Bradbury *et al.* 2006). In addition, the IBD for the two regions varied dramatically, with a much steeper slope seen in the Newfoundland samples. By comparison, the mainland IBD was more gradual and F_{ST} values seen in the Mosher's brook temporal samples correspond to mainland sites separated by ~1000 km (Bradbury *et al*. 2006). It should be noted that since the current study addresses temporal variation within a single year, it remains to be seen whether the

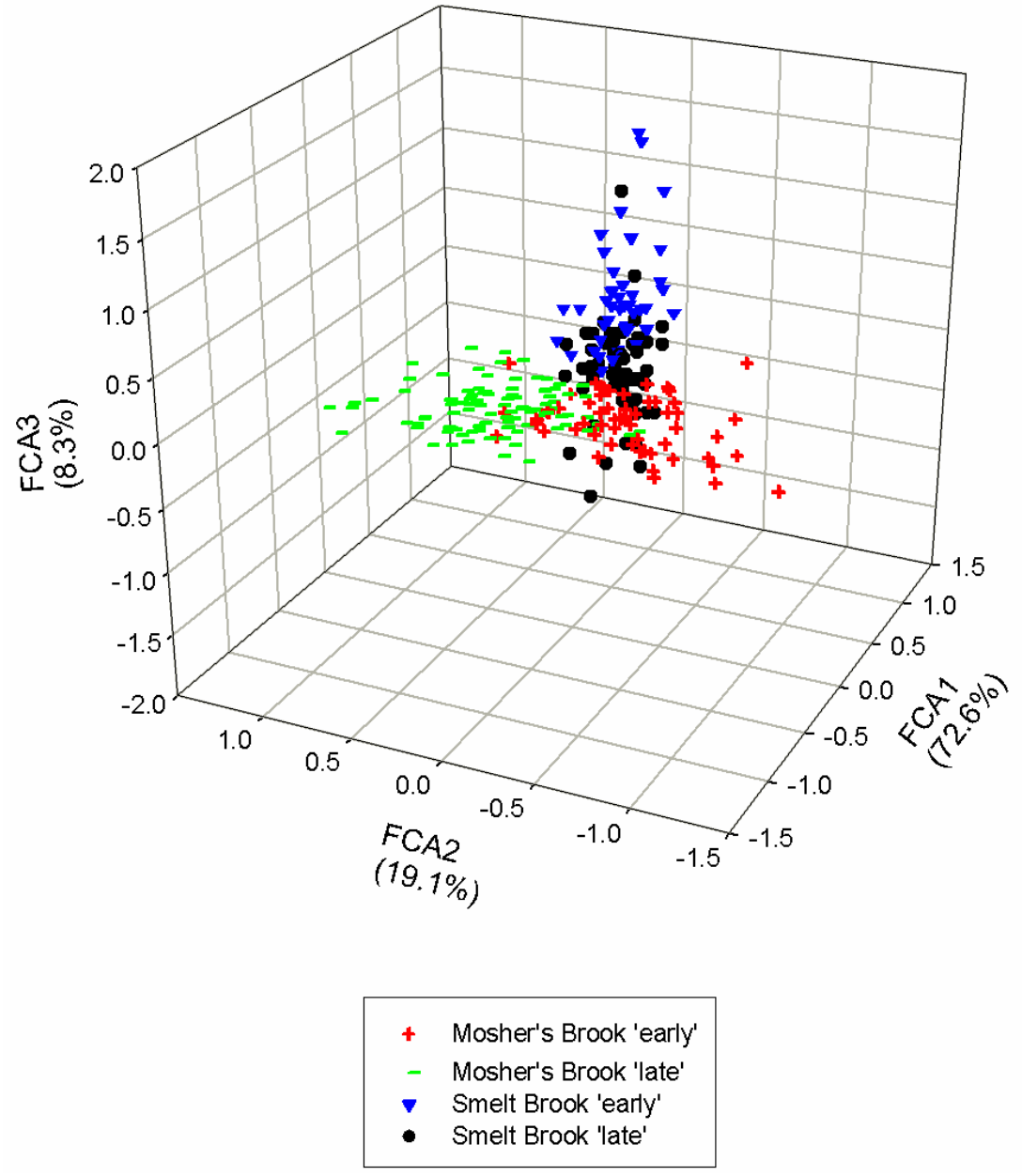

Figure 6.2 Factorial correspondence analysis of 'early' and 'late' runs for Mosher's and Smelt brooks. The first axis (separating the two locations) has been rotated to show the comparison in the temporal separation.

'early' vs. 'late' genetic differentiation remains stable across years; however, the current results are at least consistent with IBT.

6.4 DISCUSSION

The simulation results assume that that the timing of reproduction is under some degree of genetic control. It is unlikely that smelt actually spawn only on one night, or that if they do, that the choice of night is under complete genetic control. McKenzie (1964) reported that smelt in the Miramichi River spawned over 3-5 nights, but provided no information about the distribution of spawning activity over that time. The heritability of spawning time in smelt is not known, but estimates from salmonids have shown heritabilities of return time as high as 0.39 for pink salmon (*Oncorhynchus gorbuscha* Walbaum) (Smoker *et al.,* 1998) and heritabilities of spawning date in steelhead/rainbow trout (*Oncorhynchus mykiss* Walbaum) and coho salmon (*Oncorhynchus kisutch* Walbaum) have been estimated to be 0.55 to 0.57, respectively (cf. Woody *et al.*, 2000). Assuming that the heritability of spawning date is similar in smelt, a variance in spawning date on the order of several days seems plausible. The simulation model suggested that the threshold for genetic differentiation might occur with a heritable spawning period of ~3 days. Additionally, there is some evidence (McKenzie, 1964; authors, personal observations) that male smelt spawn over more nights than females. Consequently, 'temporal' gene flow might be male biased, and temporal genetic differentiation might be more apparent with maternally inherited genetic markers. It would be interesting to test this possibility by surveying early and late spawning smelt in the two streams using (maternally inherited) mitochondrial markers.

Water temperatures might be an important agent of natural selection influencing the two Mosher's Brook smelt runs. Water temperature taken on several nights during the 'early' run in Mosher's Brook was generally between 6-8^{o}C while for the 'late' run the temperature was 14-16^{o}C. This difference in temperature is expected to influence embryonic development and egg hatching rates (McKenzie, 1964), but may also have driven temperature-specific adaptations in the two spawning groups in Mosher's Brook, as has been seen in some early- and late-return populations of salmon (e.g., Brykov *et al.*, 1999). This is an area for further investigation.

The results of this study suggest that timing of reproduction in smelt can be an important factor contributing to genetic differentiation, and in this regard it provides some of the first evidence for the role of temporal differentiation in non-salmonid anadromous fishes. Indeed, the early and late spawning groups in Mosher's Brook were as genetically differentiated as populations of smelt spawning at similar times hundreds of km apart in Nova Scotia (Bradbury *et al.*, 2006). In this study, the effects of spatial isolation were deliberately eliminated by focusing on temporal isolation within spawning habitats. The majority of smelt populations, however, are separated by varying amounts of both spatial and temporal isolation, which suggests that both factors need to be considered to understand the pattern of gene flow and differentiation in this species.

CHAPTER 7

CONCLUSION

While determination of the spatial scale at which populations are genetically structured should seem routine, the underlying complexity in both a species' habitat and life-history and ecological variation makes this task all the more challenging. With advances in genetic screening and the increased use of genetic markers subject to selection (e.g., Hauser & Seeb 2008; Canino *et al*. 2005; Bradbury *et al.* 2010), the paradigm of 'open' marine populations in the absence of physical barriers and planktonic larval stages, has changed (Hauser & Carvalho 2008). With advances in genetic technology it is increasingly common to be able to apply 100s-1000s of genetic markers in non-model species lacking a reference genome. The application of next-generation sequencing coupled with advances in computation and analytical techniques is allowing researchers to increase the accuracy and efficiency of genetic applications in ecology and evolutionary biology. These advances will undoubtedly increase our understanding not only of the spatial scale of population structuring but will allow us the ability to hone in on particular traits associated with local adaptation and provide insights into the potential to deal with a changing environment.

This thesis contributed to our understanding into the ecology and evolution of rainbow smelt in particular, but highlights the importance of considering both historical and contemporary processes across the range of any organism of interest. While summarizing overall levels of genetic differentiation allows for ease of comparisons among samples, the choice of which measure to use needs consideration. To this end, I explored the use of several different measures in describing population genetic structure.

While there has been a recent debate in the literature regarding the use of various measures, it is clear that no single approach will be most suitable given the breadth of questions and systems of interest to researchers. More importantly is the choice of the most appropriate measure given the question of interest, as well as an understanding and recognition of the assumptions and limitations that will inevitably accompany any chosen method. I demonstrated that interpreting traditional F_{ST} based methods as measures of true differentiation can result in underestimating the degree of allelic differentiation among populations and species. Most notably, among fishes, this disproportionately affects marine species, which have historically been assumed to represent 'open' populations, although numerous molecular studies have begun to challenge this assumption.

While previous studies (Taylor & Bentzen 1993; Bernatchez 1997) have identified two distinct mtDNA glacial lineages with a zone of secondary contact in the upper St. Lawrence estuary (Bernatchez 1997), more comprehensive geographical coverage, particularly in Newfoundland, identified the occurrence of the 'B' mtDNA type primarily associated with the Avalon Peninsula and resulting in a transition zone to the north and west. As a result of this finding, a revision of earlier recolonization scenarios of rainbow smelt to their native range was required. The distribution of the 'B' mtDNA type in the western and eastern most locations, with the 'A' type predominating at intermediate longitudes is a striking pattern. Given a signature of a more recent population expansion of the 'A' type suggests that the 'B' type may have previously been more widespread, as they two forms are estimated to have diverged several hundred thousand years ago, well before the last glacial maximum.

Chapter 4 developed a suite of polymorphic tetranucleotide microsatellites specifically for rainbow smelt, which were used in a large-scale survey of genetic variation. Contrasting patterns of both diversity and differentiation within Newfoundland versus the mainland, suggest a more recent colonization of the Avalon Peninsula, as was also suggested by the mtDNA data. High genetic diversity at a location (for both mitochondrial and microsatelite DNA) in southwest Newfoundland was consistent with levels observed in the mainland and indeed this site clustered preferentially with mainland sites rather than Newfoundland. This suggests a possible source of colonization of western and southern Newfoundland from the 'A' mtDNA despite the presence of the Cabot Strait, a likely barrier to contemporary dispersal. Among mainland populations, despite much weaker levels of genetic structuring, regional patterns were evident in the slopes of isolation-by-distance as well as evidence for a closer relationship between the Bay of Fundy and Gulf of St. Lawrence region. Such an association may reflect a previous connection between these waterbodies during periods of higher sea levels. Overall, these results further support the roles of local retentive processes and active vertical migration of smelt, consistent with a member-vagrancy model of population genetic structure.

Chapter 7 documented the occurrence of weak, but significant genetic structure between an 'early' and 'late' run of smelt in Mosher's Brook, Musquodoboit Harbour, Nova Scotia. Genetic data and simulations confirmed the existence of a break in spawning between two peak times can lead to isolation by time (Hendry & Day 2005). As there have been other reports of bimodal run timing (MacKenzie 1964; M. Coulson, personal observations), it would be of interest to expand upon these situations as they

may be more common for smelt, especially in areas of high egg deposition. Additionally, demonstration of temporal stability in these signatures is needed to confirm differences in spawning time as a potential isolating mechanism in rainbow smelt.

Rainbow smelt offer the potential to be a model species in order to understand intraspecific variation and ultimately the processes underpinning speciation. The dichotomous nature of many aspects of their ecology and evolution allows for comparisons and contrasts among different environments, life-history strategies and historical signatures. The presence of both anadromous and landlocked forms, different morphotypes within lacustrine populations, two historical glacial races, and high versus low gene-flow environments and their associations with morphological divergence (Bradbury *et al.* 2006) offer a range of possibilities with which newer genetic tools will surely advance our understanding of the ecological and evolutionary forces at play in underpinning population structure and dynamics.

LITERATURE CITED

Aboim MA, Mavarez J, Bernatchez L, Coelho MM (2010) Introgressive hybridization between two Iberian endemic cyprinid fish: a comparison between two independent hybrid zones. Journal of Evolutionary Biology 23: 817-828.

Andoh T, Goto A (1988) Biochemical-evidence for reproductive isolation between the sympatric populations of *Cottus amblystomopsis* and *Cottus nozawae*. Japanese Journal of Ichthyology 35: 176-183.

April J, Turgeon J (2006) Phylogeography of the banded killifish (*Fundulus diaphanous*): glacial races and secondary contact. Journal of Fish Biology 69: 212-228.

Asmussen MA, Arnold J, Avise JC (1989) The effects of assortative mating and migration on cytonuclear associations in hybrid zones. Genetics 122: 923-934.

Atarhouch T, Ruber L, Gonzalez EG, Albert EM, Rami M, Dakkak A, Zardoya R (2006) Signature of an early genetic bottleneck in a population of Moroccan sardines (*Sardina pilchardus*). Molecular Phylogenetics and Evolution 39: 373-383.

Avise JC (1998) The history and purview of phylogeography: a personal reflection. Molecular Ecology 7: 371-379.

Avise JC (2000) Stability, equilibrium and molecular aspects of conservation in marine species. Hydrobiologia 420: XI-XIII.

Avise JC, Neigel JE, Arnold J (1984) Demographic influences on mitochondrial DNA lineage survivorship in animal populations. Journal of Molecular Evolution 20: 99-105.

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC (1987) Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. Annual Review of Ecology and Systematics 18: 489-522.

Baby MC, Bernatchez L, Dodson JJ (1991) Genetic structure and relationships among anadromous and landlocked populations of rainbow smelt, *Osmerus mordax*, Mitchill, as revealed by mtDNA restriction analysis. Journal of Fish Biology 39: 61-68.

Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. Molecular Ecology 13: 729-744.

Balloux F (2001) EASYPOP (version 1.7): A computer program for population genetics simulations. Journal of Heredity 92: 301-302.

Barluenga M, Sanetra M, Meyer A (2006) Genetic admixture of burbot (Teleostei: *Lota lota*) in Lake Constance from two European glacial refugia. Molecular Ecology 15:

3583-3600.

Barrette M-F, Daigle G, Dodson JJ (2009) Intraspecific vicariant history and the evolution of adaptive morphological diversity in a fish species (*Osmerus mordax*). Biological Journal of the Linnean Society 97: 140-151.

Beacham TD, Hay DE, Le KD (2005) Population structure and stock identification of eulachon (*Thaleichthys pacificus*), an anadromous smelt, in the Pacific Northwest. Marine Biotechnology 7: 363-372.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. Proceedings of the Royal Society of London Series B 263: 1619-1626.

Bekkevold D, Andre C, Dahlgren TG, Clausen LAW, Torstensen E, Mosegaard H, Carvalho GR, Christensen TB, Norlinder E, Ruzzante DE (2005) Environmental correlates of population differentiation in Atlantic herring. Evolution 59: 2656-2668.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004) Genetix 4.02, *Logiciel sous WindowsTM pour la Genetique des Populations.* Laboratoire Genome, Populations, Interactions, CNRS UPR 9060, Universite de Montpellier II, Montpellier, France.

Bermingham E, Avise JC (1986) Molecular zoogeography of freshwater fishes in the southeastern United States. Genetics 113: 939-965.

Bernatchez L (1997) Mitochondrial DNA analysis confirms the existence of two glacial races of rainbow smelt *Osmerus mordax* and their reproductive isolation in the St. Lawrence River estuary (Quebec, Canada). Molecular Ecology 6: 73-83.

Bernatchez L, Dodson JJ (1991) Phylogeographic structure in mitochondrial DNA of the lake whitefish (*Coregonus clupeaformis*) and its relation to Pleistocene glaciations. Evolution 45: 1016-1035.

Bernatchez L, Dodson (1994) Phylogenetic relationships among Palearctic and Nearctic whitefish (*Coregonus* sp.) populations as revealed by mitochondrial DNA variation. Canadian Journal of Fisheries and Aquatic Sciences 51: 240-251.

Bernatchez L, Martin S (1996) Mitochondrial DNA diversity in anadromous rainbow smelt, *Osmerus mordax* Mitchill: A genetic assessment of the member-vagrant hypothesis. Canadian Journal of Fisheries and Aquatic Sciences 53: 424-433.

Bernatchez L, Wilson CC (1998) Comparative phylogeography of nearctic and palearctic fishes. Molecular Ecology 7: 431-452.

Bohonak AJ (2002) IBD (isolation by distance): a program for analyses of isolation by distance. Journal of Heredity 93: 153-154.

Bos DH, Gopurenko D, Williams RN, DeWoody JA (2008) Inferring population history and demography using microsatellites, mitochondrial DNA, and major histocompatibility complex (MHC) genes. Evolution 62: 1458-1468.

Botsford LW, Hastings A, Gaines SD (2001) Dependence of sustainability on the configuration of marine reserves and larval dispersal distance. Ecology Letters 4: 144-150.

Bourret V, Kent MP, Primmer CR, Vasemagi A, Karlsson S, Hindar K, McGinnity P, Verspoor E, Bernatchez L, Lien S (2013) SNP-array reveals genome-wide patterns of geographical and potential adaptive divergence across the natural range of Atlantic salmon (*Salmo salar*). Molecular Ecology 22: 532-551.

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368: 455-457.

Bradbury IR, Campana SE, Bentzen P, Snelgrove PVR (2004) Synchronized hatch and its ecological significance in rainbow smelt *Osmerus mordax* in St. Mary's Bay, Newfoundland. Limnology and Oceanography 49: 2310-2315.

Bradbury IR, Coulson MW, Campana SE, Bentzen P (2006a) Morphological and genetic differentiation in anadromous smelt *Osmerus mordax* (Mitchill): disentangling the effects of geography and morphology on gene flow. Journal of Fish Biology 69: 95-114.

Bradbury IR, Gardiner K, Snelgrove PVR, Campana SE, Bentzen P, Guan L (2006b) Larval transport, vertical distribution, and localized recruitment in anadromous rainbow smelt (*Osmerus mordax*). Canadian Journal of Fisheries and Aquatic Sciences 63: 2822-2836.

Bradbury IR, Laurel B, Snelgrove PVR, Bentzen P, Campana SE (2008a) Global patterns in marine dispersal estimates: The influence of geography, taxonomic category and life history. Proceedings of the Royal Society of London Series B 275: 1803-1809.

Bradbury IR, Campana SE, Bentzen P (2008b) Low genetic connectivity in an estuarine fish with pelagic larvae. Canadian Journal of Fisheries and Aquatic Sciences 65: 147-158.

Bradbury IR, Campana SE, Bentzen P (2008c) Otolith elemental composition and adult tagging reveal spawning site fidelity and estuarine dependency in rainbow smelt. Marine Ecology Progress Series 368: 255-268.

Bradbury IR, Hubert S, Higgins B, Borza T, Bowman S, Paterson IG, Snelgrove PVR, Morris CJ, Gregory RS, Hardie DC, Hutchings JA, Ruzzante DE, Taggart CT, Bentzen P (2010) Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in

response to temperature. Proceedings of the Royal Society B – Biological Sciences 277: 3725-3734.

Bradbury IR, Coulson MW, Cook AM, Bentzen P (2010) Evidence for divergence and adaptive isolation in post-glacially derived bimodal allopatric and sympatric rainbow smelt populations. Biological Journal of the Linnean Society 101: 583-594.

Bradbury IR, Bradford R, Bentzen P (2011a) Genetic and phenotypic diversity and divergence in sympatric Lake Utopia rainbow smelt (*Osmerus mordax*). DFO Canadian Science Advisory Secretariat Research Document 2011/008: vi+28p.

Bradbury IR, Coulson MW, Campana SE, Paterson IG, Bentzen P (2011b) Contemporary nuclear and mitochondrial genetic clines in a north temperate estuarine fish reflect Pleistocene vicariance. Marine Ecology Progress Series 438: 207-218.

Brykov VA, Polyakova N, Skurikhina LA, Kukhlevsky AD, Kirillova ON, Churikov D, Pudovkin AI, Gharrett AJ (1999) Analysis of mtDNA indicates weak temporal genetic heterogeneity in pink salmon spawning runs in two rivers on Sakhalin Island. Journal of Fish Biology 55: 617-635.

Canino MF, O'Reilly PT, Hauser L, Bentzen P (2005) Genetic differentiation in walleye Pollock (*Theragra chalcogramma*) in response to selection at the pantophysin (PanI) locus. Canadian Journal of Fisheries and Aquatic Sciences 62: 2519-2529.

Canino MF, Spies IB, Cunningham KM, Hauser L, Grant WS (2010) Multiple ice-age refugia in Pacific cod, *Gadus macrocephalus*. Molecular Ecology 19: 4339-4351.

Catto NR (1998) The pattern of glaciation on the Avalon Peninsula of Newfoundland. Geographie Physique et Quaternaire 52: 23-45.

Charlesworth B (1998) Measures of divergence between populations and the effect of forces that reduce variability. Molecular Biology and Evolution 15: 538-543.

Cimmaruta R, Bondanelli P, Ruggi A, Nascetti G (2008) Genetic structure and temporal stability in the horse mackerel (*Trachurus trachurus*). Fisheries Research 89: 114-121.

CLIMAP (1976) The surface of the ice-age earth. *Science* 191: 1131-1137.

Cockerham CC, Weir BS (1987) Correlations, descent measures - drift with migration and mutation. Proceedings of the National Academy of Sciences USA 84: 8512-8514.

Cockerham CC, Weir BS (1993) Estimation of gene flow from F-statistics. Evolution 47: 855-863.

Colbeck GJ, Turgeon J, Sirois P, Dodson JJ (2011) Historical introgression and the role of selective vs. neutral processes in structuring nuclear genetic variation (AFLP) in a

circumpolar marine fish, the capelin (*Mallotus villosus*). Molecular Ecology 20: 1976-1987.

Coleman AP (1926) Pleistocene of Newfoundland. Journal of Geology 34: 193-223.

Coope GR (1979) Late Cenozoic fossil Coleoptera: Evolution, biogeography, and ecology. Annual Review of Ecology and Systematics 10: 247-267.

Copeman DG (1977) Population differences in rainbow smelt, *Osmerus mordax* multivariate analysis of mensural and meristic data. *Journal of Fisheries Research Board of Canada* 34: 1220-1229.

Corander J, Waldmann P, Sillanpaa MJ (2003) Bayesian analysis of genetic differentiation between populations. Genetics 163: 367-374.

Corander J, Marttinen P (2006) Bayesian identification of admixture events using multilocus molecular markers. Molecular Ecology 15: 2833-2843.

Coulson MW, Paterson IG, Green A, Kepkay R, Bentzen P (2006) Characterization of di- and tetranucleotide microsatellite markers in rainbow smelt (*Osmerus mordax*). Molecular Ecology Notes 6: 942-944.

Crossman EJ, McAllister DE (1986) Zoogeography of freshwater fishes of the Hudson Bay drainage, Ungava Bay and the Arctic Archipelago. In: *The Zoogeography of North American Freshwater Fishes* (eds Hocutt CH, Wiley EO), John Wiley and Sons, New York. Pages 53-104.

Curry RA (2007) Late glacial impacts on dispersal and colonization of Atlantic Canada and Maine by freshwater fishes. Quaternary Research 67: 225-233.

Curry RA, Currie SL, Bernatchez L, Saint-Laurent R (2004) The rainbow smelt, *Osmerus mordax*, complex of Lake Utopia: threatened or misunderstood? Environmental Biology of Fishes 69: 153-166.

Danzmann RG, Morgan RP, Jones MW, Bernatchez L, Ihssen PE (1998) A major sextet of mitochondrial DNA phylogenetic assemblages extant in eastern North American brook trout (*Salvelinus fontinalis*): distribution and postglacial dispersal patterns. Canadian Journal of Zoology 76: 1300-1318.

Dewoody JA, Avise JC (2000) Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. Journal of Fish Biology 56: 461-473.

Dodson JJ, Tremblay S, Colombani F, Carscadden JE, Lecomte F (2007) Trans-Arctic dispersals and the evolution of a circumpolar marine fish species complex, the capelin (*Mallotus villosus*). Molecular Ecology 16: 5030-5043.

Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. Molecular Biology and Evolution 22: 1185-1192.

Drummond A, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7: 214.

Dubois J-M, Dionne J-C (1985) The Quebec North Shore Moraine System: a major feature of Late Wisconsin deglaciation. In: *Late Pleistocene History of Northeastern New England and Adjacent Quebec* (eds. Borns HW, LaSalle P, Thompson WB). Geological Society of America, Boulder CO, USA. Pp. 125-134.

Duvernell DD, Lindmeier JB, Faust KE, Whitehead A (2008) Relative influences of historical and contemporary forces shaping the distribution of genetic variation in the Atlantic killifish, *Fundulus heteroclitus*. Molecular Ecology 17: 1344-1360.

Dyke AS, Andrews JT, Clark PU, England JH, Miller GH, Shaw J, Veillette JJ (2002) The Laurentide and Innuitian ice sheets during the Last Glacial Maximum. Quaternary Science Review 21: 9-21.

Eckert CG, Samis KE, Lougheed SC (2008) Genetic variation across species' geographical ranges: The central-marginal hypothesis and beyond. Molecular Ecology 17: 1170-1188.

Egan SP, Nosil P, Funk DJ (2008) Selection and genomic differentiation during ecological speciation: Isolating the contributions of host association via a comparative genome scan of neochlamisus bebbianae leaf beetles. Evolution 62: 1162-1181.

Elphinstone MS, Hinten GN, Anderson MJ, Nock CJ (2003) An inexpensive and high-throughput procedure to extract and purify total genomic DNA for population studies. Molecular Ecology 3: 317-320.

Evans DO, Waring P (1987) Changes in the multispecies, winter angling fishery of Lake Simcoe, Ontario, 1961-83: Invasion by rainbow smelt, *Osmerus mordax*, and the roles of intraspecific and interspecific interactions. Canadian Journal of Fisheries and Aquatic Sciences 44: 182-197.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10: 564-567.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164: 1567-
1587.

Fishback AG, Danzmann RG, Ferguson MM (2000) Microsatellite allelic heterogeneity among hatchery rainbow trout maturing in different seasons. Journal of Fish Biology 57: 1367-1380.

Francois O, Durand E (2010) Spatiall explicit Bayesian clustering models in population genetics. Molecular Ecology Resources 19: 773-784.

Frechet A, Dodson JJ, Powles H (1983a) Use of variation in biological characters for the classification of anadromous rainbow smelt (*Osmerus mordax*) groups. Canadian Joural of Fisheries and Aquatic Sciences 40: 718-727.

Frechet A, Dodson JJ, Powles H (1983b) Parasites of the anadromous American smelt (*Osmerus mordax*) of Quebec and their usefulness as biological tags. Canadian Journal of Zoology 61: 621-626.

Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147: 915.

Gagnon MC, Angers B (2006) The determinant role of temporary proglacial drainages on the genetic structure of fishes. Molecular Ecology 15: 1051-1065.

Gerlach G, Jueterbock A, Kraemer PH, Deppermann J, Harmand P (2010) Calculations of population differentiation based on G_{ST} and D: forget G_{ST} but not all of statistics! Molecular Ecology 19: 3845-3852.

Gharrett AJ, Gray AK, Brykov V (2001) Phylogeographic analysis of mitochondrial DNA variation in Alaskan coho salmon, *Oncorhynchus kisutch*. Fishery Bulletin 99: 528-544.

Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. Genetics 139: 463-471.

Goudet J (1995) FSTAT (version 1.2): A computer program to calculate F-statistics. Journal of Heredity 86: 485-486.

Gordos K, Kenchington E, Hamilton LC, Nakashima BS, Taggart CR (2005) Atlantic capelin (*Mallotus villosus*) tetranucleotide microsatellites. Molecular Ecology Notes 5: 220-222.

Goudet J (1995) FSTAT: A computer program to calculate F-statistics. Journal of Heredity 86: 485-486.

Grosberg RK, Cunningham CW (2001) Genetic structure in the sea: from populations to communities. In: *Marine Community Ecology* (eds. MD Bertness, S Gaines, ME Hay), Sinauer Associates, Sunderland, MA, pp. 61-84.

Hamilton MB, Pincus EL, Di Fiore A, Fleischer RC (1999) Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. *Biotechniques* 27: 500-507.

Hasselman DJ, Ricard D, Bentzen P (2013) Genetic diversity and differentiation in a wide ranging anadromous fish, American shad (*Alosa sapidissima*), is correlated with latitude. Molecular Ecology 22: 1558-1573.

Hauser L, Carvalho GR (2008) Paradigm shifts in marine fisheries genetics: Ugly hypotheses slain by beautiful facts. Fish and Fisheries 9: 333-362.

Hauser L, Seeb JE (2008) Advances in molecular technology and their impact on fisheries genetics. Fish and Fisheries 9: 473-486.

Hedrick PW (1999) Perspective: Highly variable loci and their interpretation in evolution and conservation. Evolution 53: 313-318.

Hedrick PW (2005) A standardized genetic differentiation measure. Evolution 59: 1633-1638.

Heller R, Siegismund HR (2009) Relationship between three measures of genetic differentiation G(ST), D-EST and G'(ST): How wrong have we been? Molecular Ecology 18: 2080-2083.

Henderson EP (1972) Surficial geology of Avalon Peninsula, Newfoundland. Geological Survey of Canada, memoir 368.

Hendry AP, Day T (2005) Population structure attributable to reproductive time: isolation by time and adaptation by time. Molecular Ecology 14: 901-916.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. Biological Journal of the Linnean Society 58: 247-276.

Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. Nature 405: 907-913.
Hutchison DW, Templeton AR (1999) Correlation between pairwise genetic and geographic distance measures: Inferring the relative influences of gene flow and drift on the distribution of genetic variability. Evolution 53: 1898-1914.

Hrabik TR, Magnuson JJ, McLain AS (1998) Predicting the effects of rainbow smelt on native fishes in small lakes: evidence from long-term research on two lakes. Canadian Journal of Fisheries and Aquatic Sciences 55: 1364-1371.

Hrabik TR, Carey MP, Webster MS (2001) Interactions between young-of-the-year exotic rainbow smelt and native yellow perch in a northern temperate lake. Transactions of the American Fisheries Society 130: 568-582.

Hubert N, Duponchelle F, Nunez J, Garcia-Davila C, Paugy D, Renno JF (2007) Phylogeography of the piranha genera *Serrasalmus* and *Pygocentrus*: implicatons for the diversification of the Neotropical ichthyofauna. Molecular Ecology 16: 2115-2136.

Hutchings JA, Gerber L (2002) Sex-biased dispersal in a salmonid fish. Proceedings of the Royal Society B – Biological Sciences 269: 2487-2493.

Hutchison DW, Templeton AR (1999) Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability.

Iles TD, Sinclair M (1982) Atlantic herring: stock discreteness and abundance. Science 215: 627-633.

Johnston CE, Cheverie JC (1988) Observations on the diel and seasonal drift of eggs and larvae of anadromous rainbow smelt, *Osmerus mordax*, and blueback herring, *Alosa aestivalis*, in a coastal stream on Prince Edward Island. Canadian Field Naturalist 102: 508-514.

Jost L (2008) G(ST) and its relatives do not measure differentiation. Molecular Ecology 17: 4015-4026.

Jost L (2009) D vs. G(ST): Response to Heller and Siegismund (2009) and Ryman and Leimar (2009). Molecular Ecology 18: 2088-2091.

Kalinowski ST (2002) Evolutionary and statistical properties of three genetic distances. Molecular Ecology 11: 1263-1273.

Kalinowski ST (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. Molecular Ecology Notes 5: 187-189.

Kalinowski ST (2011) The computer program STRUCTURE does not reliably identify the
main genetic clusters within species: simulations and implications for human population structure. Heredity 106: 625-632.

Kendall WC (1927) The smelts. Fishery Bulletin of the Fish and Wildlife Service 42: 217-325.

Kimura M (1980) Average time until fixation of a mutant allele in a finite population under continued mutation pressure: studies by analytical, numerical and pseudosampling methods. Proceedings of the National Academy of Science USA 77: 522-526.

Kittlein MJ, Gaggiotti O (2008) Interactions between environmental factors can hide isolation by distance patterns: a case study of *Ctenomys rionegrensis* in Uruguay. Proceedings of the Royal Society B – Biological Sciences 275: 2633-2638.

Knutsen H, Jorde PE, Andre C, Stenseth NC (2003) Fine-scaled geographical population structuring in a highly mobile marine species: the Atlantic cod. Molecular Ecology 12: 385-394.

Kovach AI, Breton TS, Enterline C, Berlinsky DL (2013) Identifying the spatial scale of population structure in anadromous rainbow smelt (*Osmerus mordax*). Fisheries Research 141: 95-106.

Langella O (1999) Populations 1.2.30: a population genetic software. CNRS UPR9034 (Available at http://bioinformatics.org/~tryphon/populations).

Latch EK, Dharmarajan G, Glaubitz JC, Rhodes OE (2006) Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. Conservation Genetics 7: 295-302.

Lecomte F, Dodson JJ (2004) Role of early life-history constraints and resource polymorphism in the segregation of sympatric populations of an estuarine fish. Evolutionary Ecology Research 6: 631-658.

Leis JM (1984) Larval fish dispersal and the East Pacific Barrier. Oceanographie Tropicale 19: 181-192.

Leis JM, Caselle JE, Bradbury IR, Kristiansen T, Llopiz JK, Miller MJ, O'Connor MI, Paris CB, Shanks AL, Sogard SM, Swearer SE, Treml EA, Vetter RD, Warner RR (2013) Does fish larval dispersal differ between high and low latitudes? Proceedings of the Royal Society B – Biological Sciences 280: Art. No. 20130327.

Leng L, Zhang D-X (2011) Measuring population differentiation using *G*ST or *D*? A simulation study with microsatellite DNA markers under a finite island model and nonequilibrium conditions. Molecular Ecology 20: 2494-2509.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451-1452.

Liu JX, Gao TX, Wu SF, Zhang YP (2007) Pleistocene isolation in the Northwestern Pacific marginal seas and limited dispersal in a marine fish, *Chelon haematoceheilus* (Temminck & Schlegel, 1845). Molecular Ecology 16: 275-288.

Liu JX, Tatarenkov A, Beacham TD, Gorbachev V, Wildes S, Avise JC (2011) Effects of Pleistocene climatic fluctuations on the phylogeographic and demographic histories of Pacific herring (*Clupea pallasii*). Molecular Ecology 20: 3879-3893.

Lotterhos KE (2012) Nonsignificant isolation by distance implies limited dispersal. Molecular Ecology 21: 5637-5639.

Maggs CA, Castilho R, Foltz D, Henzler C, Jolly MT, Kelly J, Olsen J, Perez KE, Stam W, Vainola R, Viard F, Wares J (2008) Evaluating signatures of glacial refugia for North Atlantic benthic marine taxa. Ecology 89: 108-122.

Manica A, Prugnolle F, Balloux F (2005) Geography is a better determinant of human genetic differentiation than ethnicity. Human Genetics 118: 366-371.

Maruyama T (1970) Effective number of alleles in a subdivided population. Theoretical Population Biology 23: 216-240.

McAllister DE, Platania SP, Schueler FW, Baldwin ME, Lee DS (1986) Ichthyofaunal patterns on a geographic grid. In: *The Zoogeography of North American Freshwater Fishes* (eds Hocutt CH, Wiley EO). John Wiley and Sons, New York. Pp. 17-51.

McCusker MR, Parkinson E, Taylor EB (2000) Mitochondrial DNA variation in rainbow trout (*Oncorhynchus mykiss*) across its native range: testing biogeographical hypotheses and their relevance to conservation. Molecular Ecology 9: 2089-2108.

McCusker MR, Bentzen P (2010) Historical influences dominate the population genetic structure of a sedentary marine fish, Atlantic wolffish (*Anarhichas lupus*), across the North Atlantic Ocean. Molecular Ecology 19: 4228-4241.

McKenzie RA (1964) Smelt life history and fishery in the Miramichi River, New Brunswick. Fisheries Research Board of Canada Bulletin 114.

McLean JE, Taylor EB (2001) Resolution of population structure in an species with high gene flow: microsatellite variation in the eulachon (Osmeridae: *Thaleichthys pacificus*). Marine Biology 139: 411-420.

McPhail JD, Lindsey CC (1970) Freshwater fishes of northwestern Canada and Alaska. Fisheries Research Board of Canada, Bulletin No. 173.

Meirmans PG (2006) Using the AMOVA framework to estimate a standardized genetic differentiation measure. Evolution 60: 2399-2402.

Meirmans PG, Van Tienderen PH (2004) GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. Molecular Ecology Notes 4: 792-794.

Meirmans PG, Hedrick PW (2010) Assessing population structure: F_{ST} and related measures. Molecular Ecology Resources 11: 5-18.

Mercado-Silva N, Olden JD, Maxted J, Hrabik TR, Vander Zanden MJ (2006) Forecasting the spread of invasive rainbow smelt in the Laurentian Great Lakes region of North America. Conservation Biology 20: 1740-1749.

Mills LS, Allendorf FW (1996) The one-migrant-per-generation rule in conservation and management. Conservation Biology 10: 1509-1518.

Moore J-S (2012) Patterns and consequences of dispersal in Arctic char (Pisces: *Salvelinus alpinus*) from the Canadian Arctic. PhD Thesis, University of British Columbia, British Columbia, Canada. 196 pages.

Nagylaki T (1998) Fixation indices in subdivided populations. Genetics 148: 1325-1332.

Narum SR (2006) Beyond Bonferroni: Less conservative analyses for conservation genetics. Conservation Genetics 7: 783-787.

Nauta MJ, Weissing FJ (1996) Constraints on allele size at microsatellite loci: Implications for genetic differentiation. Genetics 143: 1021-1032.

Nei M (1973) Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences USA 70: 3321-3323.

Nei M (1975) Molecular population genetics and evolution. North-Holland, New York.

Neigel JE (2002) Is F_{ST} obsolete? Conservation Genetics 3: 167-173.

Nellbring S (1989) The ecology of smelts (Genus *Osmerus*): a literature review. Nordic Journal of Freshwater Research 65: 116-145.

Nielsen EE Nielsen PH, Meldrup D, Hansen MM (2004) Genetic population structure of turbot (*Scophthalmus maximus* L.) supports the presence of multiple hybrid zones for marine fishes in the transition zone between the Baltic Sea and the North Sea. Molecular Ecology 13: 585-595.

Olsen EM, Knutsen H, Gjosaeter J, Jorde PE, Knutsen JA, Stenseth NC (2008) Small-scale biocomplexity in coastal Atlantic cod supporting a Darwinian perspective on fisheries management. Evolutionary Applications 1: 524-533.

Orsini L, Corander J, Alasentie A, Hanski I (2008) Genetic spatial structure in a butterfly metapopulation correlates better with past than present demographic structure. Molecular Ecology 17: 2629-2642.

Ouellet P, Dodson JJ (1985) Dispersion and retention of anadromous rainbow smelt (*Osmerus mordax*) larvae in the middle estuary of the St. Lawrence River. Canadian Journal of Fisheries and Aquatic Sciences 42: 332-341.

Paetkau D, Waits LP, Clarkson PL, Craighead L, Strobeck C (1997) An empirical evaluation of genetic distance statistics using microsatellite data from bear (Ursidae) populations. Genetics 147: 1943-1957.

Palstra FP, O'Connell MF, Ruzzante DE (2007) Population structure and gene flow reversals in Atlantic salmon (*Salmo salar*) over contemporary and long-term temporal scales: effects of population size and life history. Molecular Ecology 16: 4504-4522.

Palumbi SR (2003) Population genetics, demographic connectivity, and the design of marine reserves. Ecological Applications 13: S146-S158.

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6: 288-295.

Peakall R, Smouse PE (2012) GENALEX 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. Bioinformatics 28: 2537-2539.

Petersen CH (2007) Historical demography and contemporary spatial genetic structure of an estuarine crab in the northeast Pacific (*Hemigrapsus oregonensis*). Marine Biology 150: 1289-1300.

Pielou EC (1991) After the ice age: The return of life to glaciated North America. University of Chicago Press, Chicago, USA. ix+366 pages.

Pigeon D, Dodson JJ, Bernatchez L (1998) A mtDNA analysis of spatiotemporal distribution of two sympatric larval populations of rainbow smelt (*Osmerus mordax*) in the St. Lawrence River estuary, Quebec, Canada. Canadian Journal of Fisheries and Aquatic Sciences 55: 1739-1747.

Piry S, Alapetite A, Cornuet J-M, Paetkau D, Baudouin L, Estoup A (2004) Geneclass2: A software for genetic assignment and first-generation migrant detection. Journal of Heredity 95: 536-539.

Pogson GH (2001) Nucleotide polymorphism and natural selection at the pantophysin (Pan I) locus in the Atlantic cod, *Gadus morhua* (L.). Genetics 157: 317-330.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

Puebla O, Bermingham E, McMillan WO (2012) On the spatial scale of dispersal in coral reef fishes. *Molecular Ecology* 21: 5675-5688.

Raymond M, Rousset F (1995) Genepop (version 2.3): Population genetics software for exact tests and ecumenicism. Journal of Heredity 86: 248-249.

Rey O, Turgeon J (2007) Influence of historical events and contemporary estuarine circulation on the genetic structure of the banded killifish (*Fundulus diaphanous*) in the St. Lawrence River (Quebec, Canada). Canadian Journal of Zoology 85: 891-901.

Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. Molecular Biology and Evolution 9: 552-569.

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. Molecular Ecology Notes 4: 137-138.

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics 145: 1219-1228.

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (eds Krawetz S, Misener S), pp. 365-386. Humana Press, Totowa, New Jersey.

Rupp RS, Redmond MA (1966) Transfer studies of ecologic and genetic variation in the American smelt. Ecology 47: 253-259.

Ruzzante DE (1998) A comparison of several measures of genetic distance and population structure with microsatellite data: Bias and sampling variance. Canadian Journal of Fisheries and Aquatic Sciences 55:1-14.

Ruzzante DE, Mariani S, Bekkevold D, Andre C, Mosegaard H, Clausen LAW, Dahlgren T, Hutchinson WF, Hatfield EMC, Torstensen E, Brigham J, Simmonds EJ, Laikre L, Larsson LC, Stet RJM, Ryman N, Carvalho GR (2006) Biocomplexity in a highly migratory pelagic marine fish. Proceedings of the Royal Society B – Biological Sciences 273: 1459-1464.

Ryman N, Leimar O (2008) Effect of mutation on genetic differentiation among nonequilibrium populations. Evolution 62: 2250-2259.

Ryman N, Leimar O (2009) G(ST) is still a useful measure of genetic differentiation - a comment on Jost's D. Molecular Ecology 18: 2084-2087.

Saint-Laurent R, Legault M, Bernatchez L (2003) Divergent selection maintains adaptive differentiation despite high gene flow between sympatric rainbow smelt ecotypes (*Osmerus mordax* Mitchill). Molecular Ecology 12: 315-330.

Scott WB, Scott MG (1988) Atlantic Fishes of Canada. University of Toronto Press, Toronto, Canada, 731 pages.

Shaw J, Gareau P, Courtney RC (2002) Palaeogeography of Atlantic Canada 13-0 ka. Quaternary Science Reviews 21: 1863-1880.

Shaw J, Piper DJW, Fader GBJ, King EL, Todd BJ, Bell T, Batterson MJ, Liverman DGE (2006) A conceptual model of the deglaciation of Atlantic Canada. Quaternary Science Reviews 25: 2059-2081.

Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. Evolution 47: 264-279.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139: 457-462.

Smith GR (1992) Introgression in fishes: significance for paleontology, cladistics, and evolutionary rates. Systematic Biology 41: 41-57.

Smoker WW, Gharrett AJ, Stekoll MS (1998) Genetic variation of return date in a population of pink salmon: a consequence of fluctuating environment and dispersive selection? Alaska Fishery Research Bulletin 5: 46-54.

Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS (2006) Comparative phylogeography of unglaciated eastern North America. Molecular Ecology 15: 4261-4293.

Strange RM, Burr BM (1997) Intraspecific phylogeography of North American highland fishes: A test of the Pleistocene vicariance hypothesis. Evolution 51: 885-897.

Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF (1998) Comparative phylogeography and postglacial colonization routes in Europe. Molecular Ecology 7: 453-464.

Tajima F (1989) Statistical methods for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585.

Takahata N (1983) Gene identity and genetic differentiation of populations in the finite island model. Genetics 104: 497-512.

Taylor EB, Bentzen P (1993a) Evidence for multiple origins and sympatric divergence of trophic ecotypes of smelt (*Osmerus*) in northeastern North America. Evolution 47: 813-832.

Taylor EB, Bentzen P (1993b) Molecular genetic evidence for reproductive isolation between sympatric populations of smelt *Osmerus* in Lake Utopia, south-western New Brunswick, Canada. Molecular Ecology 2: 345-357.

Toews DPL, Brelsford A (2012) The biogeography of mitochondrial and nuclear discordance in animals. Molecular Ecology 21: 3907-3930.

Turgeon J, Bernatchez L (2001) Mitochondrial DNA phylogeography of lake cisco (*Coregonus artedi*): evidence supporting extensive secondary contacts between two glacial races. Molecular Ecology 10: 987-1001.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) Microchecker:

software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes 4: 535-538.

Verspoor (2005) Regional differentiation of North American Atlantic salmon at allozyme loci. Journal of Fish Biology 67: 80-103.

Waples RS (1998) Separating the wheat from the chaff: Patterns of genetic differentiation in high gene flow species. Journal of Heredity 89: 438-450.

Waples RS, Gaggiotti O (2006) What is a population? an empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. Molecular Ecology 15: 1419-1439.

Ward RD, Woodwark M, Skibinski DOF (1994) A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes. Journal of Fish Biology 44: 213-232.

Wares JP, Cunningham CW (2001) Phylogeography and historical ecology of the North Atlantic intertidal. Evolution 55: 2455-2469.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38: 1358-1370.

Whitlock MC (1992) Temporal fluctuations in demographic parameters and the genetic variance among populations. Evolution 46: 608-615.

Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: FST does not equal 1/(4Nm+1). Heredity 82: 117-125.

Whitlock MC (2011) G'_{ST} and D do not replace F_{ST}. Molecular Ecology 20: 1083-1091.

Wilson CC, Hebert PDN, Reist JD, Dempson JB (1996) Phylogeography and postglacial dispersal of Arctic charr *Salvelinus alpinus* in North America. Molecular Ecology 5: 187-197.

Wilson CC, Hebert PDN (1996) Phylogeographic origins of lake trout (*Salvelinus namaycush*) in eastern North America. Canadian Journal of Fisheries and Aquatic Sciences 53: 2764-2775.

Wilson CC, Hebert PDN (1998) Phylogeography and postglacial dispersal of lake trout (*Salvelinus namaycush*) in North America. Canadian Journal of Fisheries and Aquatic Sciences 55: 1010-1024.

Woody CA, Olsen JB, Reynolds J, Bentzen P (2000) Temporal variation in phenotypic and genotypic traits in two sockeye salmon populations, Tustumena Lake, Alaska. Transactions of the American Fisheries Society 129: 1031-1043.

Wright S (1943) Isolation by distance. Genetics 28: 114-138.

Wright S (1946) Isolation by distance under diverse systems of mating. Genetics 31: 39-59.

Wright S (1949) The genetical structure of populations. Annals of Human Genetics 15: 323-354.

Zane L, Marcato S, Bargelloni L, Bortolotto E, Papetti C, Simonato M, Varotto V, Patarnello T (2006) Demographic history and population structure of the Antarctic silverfish *Pleuragramma antarcticum*. Molecular Ecology 15: 4499-4511.